

Sat Jun 8 10:14:10 2002

us-09-746-783-18.rng

SEQ ID NO: 18
Database: N-Geneseq-032802
Ac. NO: AAV99731

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:17:00 ; Search time 272.98 Seconds
(without alignments)
3025.261 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481
Sequence: 1 GGATCTGTATTAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	481	100.0	481	20	AAV99731
2	289.6	60.2	318	20	AAV88049
3	70.4	14.6	631	21	AAAI6169
4	61.6	12.8	382	23	AAAD3536
5	61.6	12.8	516	21	AAAC0362
6	60.6	12.6	581	10	AAAG3083
7	60.6	12.6	581	20	AAAG0544
8	60	12.5	557	21	AAAG9778
9	60	12.5	594	21	AAAG9777

ALIGNMENTS

RESULT	1	ALIGNMENTS
AAV99731		
ID	AAV99731 standard; cDNA; 481 BP.	
AC	AAV99731:	
DT	26-APR-1999 (first entry)	
DE	Human adult testis secreted protein fg505_4 cDNA.	
XX	Human adult testis secreted protein fg505_4; ds.	
KM	Secreted protein; human; testis; fg505_4; ds.	
XX	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	84..407
FT		/tag= a
XX	WO9856909-A2.	
PN		
PD	17-DEC-1998.	
XX		
PF	08-JUN-1998; 98MO-US11822.	
PR	05-JUN-1998; 98US-0092722.	
PR	11-JUN-1997; 97US-0873218.	
XX		
PA	(GEMV) GENETICS INST INC.	
PI	Agostino MT, Fechtel K, Howes SH, Jacobs K, Lavallie ER,	
PI	MCCoy JM, Racie LA, Spaulding V, Treacy M;	
XX		

Human adult testis secreted protein fg505_4 cDNA.

Secreted protein; human; testis; fg505_4; ds.

Homo sapiens.

Key Location/Qualifiers

FT CDS 84..407

FT CDS 84..407

FT CDS 84..407

FT CDS 84..407

FT CDS 84..407

FT CDS 84..407

FT CDS 84..407

FT CDS 84..407

DR WPI: 1999-080899/07.
DR P-PSDB; AAM95351.

PT New polynucleotides encoding secreted human proteins - derived from
PT human foetal brain, adult testes, foetal kidney, adult thyroid or
PT adult retina cDNA libraries

PS Claim 30; Page 85; 113pp; English.

XX This is the nucleotide sequence of cDNA clone fg505_4, which
CC includes an open reading frame for a 107-amino acid polypeptide
CC (see AAM95351). The clone was isolated from a human adult testis
CC cDNA library using methods which are selective for cDNAs encoding
CC secreted proteins, or was identified as encoding a secreted or
CC transmembrane protein on the basis of computer analysis of the
CC amino acid sequence of the encoding protein. Database searches
CC indicate some sequence similarity to known sequences. The
CC invention provides cDNA clones (see AAV9721-33) from human adult
CC thyroid, adult retina, adult testis, foetal kidney and foetal brain
CC that encode novel secreted proteins (see AAM95344-53). Each clone is
CC individually available from deposit clone ATCC 98451 (see also
CC AAV9734-43). The isolated polynucleotides (PNS) and proteins are
CC predicted to have activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans
CC and animals, although no supporting data is given. Suggested
CC activities include nutritional, cytokine, cell proliferation or
CC differentiation, immune stimulating (e.g. as vaccines) or immune
CC suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokinetic, haemostatic,
CC thrombolytic, receptor/ligand, anti-inflammatory, cadherin/tumour
CC invasion suppressor, and tumour inhibition activities. The PNS are
CC also stated to be useful for gene therapy.

CC Sequence 481 BP; 185 A; 85 C; 90 G; 121 T; 0 other;

Query Match Best Local Similarity 100.0%; Score 481; DB 20; Length 481;
Matches 481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATACGTAAATAATAGAGACAGCTACAGTATCACTAATCAAGAGGATTTTC 60
DB 1 ggaatcgtataataatagagacagctacagtatcaactcaacacaggggttttc 60
QY 61 ATACGACATTCCTCGGTATATCATGATGATGATTAATTAAGACGAAATTAATA 120
DB 61 atcagacattccctcgggtatcatcagatcagattataaagacgaatgaattaaa 120
QY 121 CATTTTACAGCTGCGGACACAAACATCGAGTGTCAATTTCTTCGAAACGGTGTG 180
DB 121 cattttgacagctgcgcgacacaaacatcgagtggtcattttcttcgaaacgggtg 180
QY 181 GTCCCTGAAAGAGATGTTTCTGTTTCCATGAGCTGCGTGAACCTGTGCATCAATA 240
DB 181 gtccctgaaagagatgtttctgtttccatgagctgcgaactgtcacatcaaaa 240
QY 241 CAATACCCACATTTAGATGTTTCAAGAAAGCCAGAGTACCTATTCACGAATCA 300
DB 241 caataccacatltcagatgttcaagaaagccagaaggtacatccatltccaagaatca 300
QY 301 AAAGAAATTAATTTGCTGTATAGAAAGTGAATTCATGAGCAACCTGATTTTGTG 360
DB 301 aaagaataatttgcgtatagaagtgtatcattcagacacatgttttgagtttggtg 360
QY 361 GAGCCGATGCTAAATAATTTGGAGCCAAAGACTCAAGATTAATGTAACCTGATCTCAAG 420
DB 361 gagccgatgcttaataatTTGGAGCCAAAGACTCAAGATTAATGTAACCTGATCTCAAG 420
QY 421 GCAAAATGCTACTTGTGATATTGAAAGCAAAAAAATAAAAAAATAAAAAAATAAAA 480
DB 421 gcaaaatgctacttGTGATATTGAAAGCAAAAAAATAAAAAAATAAAAAAATAAAA 480
QY 481 A 481

DB 481 a 481

RESULT 2

AAV88049 ID AAV88049 standard; cDNA; 318 BP.

AAV88049;

12-FEB-1999 (first entry)

EST clone FQ505.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytic;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

OS Homo sapiens.

PN W09845437-A2.

PD 15-OCT-1998.

PP 10-APR-1998; 98WO-US06956.

PR 10-APR-1997; 97US-0837312.

XX (GEMY) GENETICS INST INC.

PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

PT Racine LA, Spaulding V, Treacy M;

DR WPI: 1999-070078/06.

PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries

PS Claim 1; Page 265; 641pp; English.

XX The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.

XX Sequence 318 BP; 99 A; 64 C; 67 G; 88 T; 0 other;

Query Match Best Local Similarity 60.2%; Score 289.6; DB 20; Length 318;
Matches 292; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 149 CGCAGGTGTTCAATTTCTGTAACGCTGTGCTCCGCAAAAGATGTTTCTGTTT 208
DB 149 cgcaggtgttcaatTTCTGTAACGCTGTGCTCCGCAAAAGATGTTTCTGTTT 208
QY 209 CCAATGAGTGGCTGAAATTTCTGCAATTAATAACATACCCATTTTGAGATTTCAAGAA 268
DB 209 ccaatgagtggctGAAATTTCTGCAATTAATAACATACCCATTTTGAGATTTCAAGAA 268
QY 269 AAGCCAGAAAGTAACCCATATCTCAAGAAATCAAAAAAATAATTTGCTGTATAGAAAGTGC 328
DB 269 aagccagaaagTAACCCATATCTCAAGAAATCAAAAAAATAATTTGCTGTATAGAAAGTGC 328

CC Immunassay. The TC-CSF coding sequence, or portions of it, may be
CC used in hybridisation assay or as primers.

XX Sequence 630 BP; 203 A; 102 C; 129 G; 196 T; 0 other:

Query Match 12.5%; Score 60; DB 14; Length 630;

Best Local Similarity 53.9%; Pred. No. 0.00021;
Matches 193; Conservative 0; Mismatches 155; Indels 10; Gaps 3;

```
QY 62 TCAGCAGCTCCCTGGTATATCATGATGATTTAAAGACAGCAATGATTTAAAC 121
    || || || || || || || || || || || || || || || || || || || ||
DB 204 tcgcagactccagccaagatggtgagcagatcgagagcagactgctttcagga 263
QY 122 ATTTTACAGCTGCGGACAGCAACGAGTGTCAATTTCTTGAAACGCTGG 181
    || || || || || || || || || || || || || || || || || || || ||
DB 264 agccttgagcgtcagagtgataaactgtagtagtactccagccacagtggtg 323
QY 182 TCCCTGCAAAAGATGTTCTGTTTCATGAGCTGGTGAACCTGTCAATCAAAAC 241
    || || || || || || || || || || || || || || || || || || || ||
DB 324 gacctgcaaaatgataagccttcttcattccctctcgtgaaagatctcaacgtgat 383
QY 242 AATACCCATTTTACAGATGTTCAAGAAAGCCAGAGGTAACCTTATTTCAAGATCAA 301
    || || || || || || || || || || || || || || || || || || || ||
DB 384 atccttgaagtagatgtg--gatgactgtcagagatgtgtctcagagtgtagaagtc 440
QY 302 AAGAAATATTTGCTGTATAGAGTGAATGATGAGCAACCTGATTTT--TGAGTTTGT 359
    || || || || || || || || || || || || || || || || || || || ||
DB 441 atgca-----tgcacaactgcaagcttctttaaagaagagcaaaagtgaggatcttct 495
QY 360 GGAGCCGATGCTAAATAATTGAGCCAGACATCAAGATTAATTAATGATGATCTCC 417
    || || || || || || || || || || || || || || || || || || || ||
DB 496 ggaagcaataagaaagcttgaagcaccattatgaattagctcaatcatgttttc 553
```

RESULT 13

AAD09895 standard; cDNA; 630 BP.

XX AAD09895;

DT 12-SEP-2001 (first entry)

DE Human T cell-derived colony stimulating factor (TC-CSF) cDNA.

KX T cell-derived colony stimulating factor; TC-CSF; immunostimulant;

KW Acquired immune deficiency syndrome; AIDS; cancer; gene therapy;

KX haematopoietic growth factor; human; ss.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 1..630

FT /tag- a

FT /product- "Human TC-CSF protein"

FT /transl_except- (pos:541..543, aa:Xaa)

FT /transl_except- (pos:562..564, aa:Xaa)

FT /transl_except- (pos:628..630, aa:Xaa)

FT /note- "Xaa corresponds to in-frame stop codon;

FT CDS does not include start and stop codon"

FT sig_peptide 55..84

FT mat_peptide 85..627

FT /tag- b

FT /product- "Mature human TC-CSF protein"

PR 23-MAY-1989; 89US-0356006.
PR 12-JUL-1991; 91US-0729135.
PR 19-AUG-1991; 91US-0747784.
PR 01-NOV-1991; 91US-0788115.
PR 01-JUL-1992; 92US-0906866.

XX (CHOU/) CHOUDHURY C.

XX Choudhury C;

DR WPI; 2001-432041/46.

DR P-PSDB; AAE05136.

PT New T cell-derived colony stimulating factor for treating

PT immune-compromised patients, including acquired immunodeficiency

PT syndrome patients and certain types of cancer patients

PS Example 20; Column 47-50; 38pp; English.

XX The invention relates to T-cell derived colony stimulating factors

CC (TC-CSF) and nucleic acid molecules encoding them. TC-CSF which

CC is a novel haematopoietic growth factor is useful for treating

CC immune-compromised patients, including acquired immune deficiency

CC syndrome (AIDS) patients and certain types of cancer patients.

CC TC-CSF DNA is used in gene therapy. The present sequence is human

XX TC-CSF cDNA.

XX Sequence 630 BP; 202 A; 110 C; 123 G; 195 T; 0 other:

Query Match 12.5%; Score 60; DB 22; Length 630;

Best Local Similarity 52.0%; Pred. No. 0.00021;
Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

```
QY 62 TCAGCAGCTCCCTGGTATATCATGATGATTTAAAGACAGCAATGATTTAAAC 121
    || || || || || || || || || || || || || || || || || || || ||
DB 204 tcgcagactccagccaagatggtgagcagatcgagagcagactgctttcagga 263
QY 122 ATTTTACAGCTGCGGACAGCAACGAGTGTCAATTTCTTGAAACGCTGG 181
    || || || || || || || || || || || || || || || || || || || ||
DB 264 agccttgagcgtcagagtgataaactgtagtagtactccagccacagtggtg 323
QY 182 TCCCTGCAAAAGATGTTCTGTTTCATGAGCTGGTGAACCTGTCAATCAAAAC 241
    || || || || || || || || || || || || || || || || || || || ||
DB 324 gacctgcaaaatgataagccttcttcattccctctcgtgaaagatctcaacgtgat 383
QY 242 AATACCCATTTTACAGATGTTCAAGAAAGCCAGAGGTAACCTTATTTCAAGATCAA 301
    || || || || || || || || || || || || || || || || || || || ||
DB 384 atccttgaagtagatgtg--gatgactgtcagagatgtgtctcagagtgtagaagtc 440
QY 302 AAGAAATATTTGCTGTATAGAGTGAATGATGAGCAACCTGATTTT--TGAGTTTGT 361
    || || || || || || || || || || || || || || || || || || || ||
DB 441 atgcatgccaacatccagcttctttaaag--aagggcaaaagtgaggatcttcg 497
QY 362 AGCCGATGCTAAATAATTGAGCCAGACATCAAGATTAATTAATGATGATCTCC 417
    || || || || || || || || || || || || || || || || || || || ||
DB 498 agccaataagaaagcttgaagcaccattatgaattagctcaatcatgttttc 553
```

RESULT 14

AAQ49402 standard; DNA; 456 BP.

XX AAQ49402;

DT 04-MAY-1994 (first entry)

DE 14kDa ECEF/thioredoxin/ADF coding sequence.

KX Eosinophil cytotoxicity enhancing factor; ECEF; stimulation;

KW diagnosis; therapy; thioredoxin; ss.

OS Homo sapiens.

Sat Jun 8 10:14:10 2002

us-09-746-783-18.rng

Page 11

Search completed: June 8, 2002, 03:25:51
job time: 4131 sec

PD 6/5/98

p 142

Sat Jun 8 10:14:12 2002

us-09-746-783-19.rag

SEARCH ID NO: 19
Database: A-Geneseg-032802
AC NO: AA095351

File Copy
Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:20:36 ; Search time 61.07 Seconds
(without alignments)
194.611 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562
Sequence: 1 MWQIKDYNFEKFLTAAGH.....IFECGADAKKLEANTQEIOM 107

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

A-Geneseg-032802: *
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	562	100.0	107	20	AA095351 Human adult testis
2	190	33.8	105	10	AA092141 Recombinant human
3	190	33.8	105	14	AA092141 Recombinant human
4	190	33.8	105	16	AA092141 Recombinant human
5	190	33.8	105	20	AA092141 Recombinant human
6	190	33.8	105	21	AA092141 Recombinant human
7	190	33.8	105	22	AA092141 Recombinant human
8	189	33.6	105	15	AA092141 Recombinant human
9	187	33.3	105	14	AA092141 Recombinant human
10	185	32.9	104	15	AA092141 Recombinant human
11	185	32.9	104	16	AA092141 Recombinant human

Result	ID	Score	Query Match	Length	Description
1	AA095351	562	100.0	107	Human adult testis secreted protein fg505_4.
2	AA095351	190	33.8	105	Human adult testis secreted protein fg505_4.
3	AA095351	190	33.8	105	Human adult testis secreted protein fg505_4.
4	AA095351	190	33.8	105	Human adult testis secreted protein fg505_4.
5	AA095351	190	33.8	105	Human adult testis secreted protein fg505_4.
6	AA095351	190	33.8	105	Human adult testis secreted protein fg505_4.
7	AA095351	190	33.8	105	Human adult testis secreted protein fg505_4.
8	AA095351	189	33.6	105	Human adult testis secreted protein fg505_4.
9	AA095351	187	33.3	105	Human adult testis secreted protein fg505_4.
10	AA095351	185	32.9	104	Human adult testis secreted protein fg505_4.
11	AA095351	185	32.9	104	Human adult testis secreted protein fg505_4.

XX Claim 31; Page 86; 113pp; English.
 PS
 CC This is the amino acid sequence of fg505.4, a novel human
 CC secreted protein predicted from the nucleotide sequence of a
 CC human adult testis cDNA clone (see AAV99731). Database searches
 CC indicate some sequence similarity to known sequences. The
 CC invention provides cDNA clones (see AAV99721-33) from human adult
 CC thyroid, adult retina, adult testis, foetal kidney and foetal
 CC brain that encode novel secreted proteins (see AAV95344-53). The
 CC polynucleotides and proteins are predicted to have activities
 CC which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals, although no
 CC supporting data is given. Suggested activities include nutritional,
 CC cytokine, cell proliferation/differentiation, immune stimulating
 CC (e.g. as vaccines) or immune suppressing, haematopoietic regulating,
 CC tissue growth, activin/inhibin, chemotactic/chemokinetic,
 CC haemostatic, thrombolytic, receptor/ligand, anti-inflammatory,
 CC cadherin/tumour invasion suppressor, and tumour inhibition
 CC activities.
 CC
 CC Sequence 107 AA:

Query Match 100.0%; Score 562; DB 20; Length 107;
 Best Local Similarity 100.0%; Pred. No. 6.2e-64;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVOIITDNEFKETFLTAAGHKLAVQSSRCGCPKRMPEVHELAETCHIKTIPFQMF 60
 DB 1 mvoiltdnefkettlaagkklavqssrcgcpkrmfpvheleatciktptqmf 60
 OY 61 KKSQKVTLSFRIRIICCYRSGFMSNLIFECGADAKKLEAKTOELM 107
 DB 61 kksqkvtlfsrirkriiccyrgsfmsnlifecgadakkkleakqelml 107

RESULT 2
 ID AAP92141
 XX AAP92141 standard; protein; 105 AA.

AC AAP92141;
 XX
 XX 25-JUN-1990 (first entry)
 DE Recombinant human adult T cell leukaemia derived factor.
 XX
 XX Human adult T cell leukaemia derived factor; hADF; cancer;
 KW Immunodeficiency disease.
 XX
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT Region 2..102
 FT /note="Claim 1"

EP299206-A.
 XX
 PD 18-JAN-1989.
 XX
 PF 10-JUN-1988; 88EP-0109311.
 XX
 PR 12-JUN-1987; 87JP-0146348.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 PI Yodol J, Tagaya Y, Maeda M, Matsui H, Kondo N, Hamuro J;
 DR WPI; 1989-016762/03.
 DR N-PSDB; AAN93083.

Recombinant human adult T cell leukaemia derived factor polypeptide -
 used for treating cancer, immunodeficiency disease etc.

XX Disclosure; 24pp; English.
 PS
 CC DNA encoding the polypeptide was sequenced from a gene bank prep. from
 CC mRNA isolated from ATL-2 cells from patient with adult T leukaemia
 CC virus. Vectors contg. the DNA can be used to transform host cells for
 CC prodn. of hADF polypeptide. The polypeptide causes differentiation and
 CC induces growth of lymphocytes and fibroblasts. The N-terminal Met is
 CC optional. Similar peptides with deletions, sustns. and allelic derivs.
 CC may also be used.
 CC See also AAP94809.
 CC
 CC Sequence 105 AA:

Query Match 33.8%; Score 190; DB 10; Length 105;
 Best Local Similarity 36.2%; Pred. No. 2.1e-16;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

OY 1 MVOIITDNEFKETFLTAAGHKLAVQSSRCGCPKRMPEVHELAETCHIKTIPFQMF 47
 DB 1 mvoiltdnefkettlaagkklavqssrcgcpkrmfpvheleatciktptqmf 60
 OY 48 -----TCHIKTIPFQMFKKSQKVTLSFRIRIICCYRSGFMSNLIFECGADAKKLE 100
 DB 61 dcqdvaseevckmptfifkkyqkv-----efsgankexle 98
 OY 101 AKTOELM 107
 DB 99 atlineiv 105

RESULT 3
 ID AAR37700
 XX AAR37700 standard; protein; 105 AA.

AC AAR37700;
 XX
 XX 09-NOV-1993 (first entry)
 DE ADF.
 XX
 XX Human; ADF; transgenic; mouse; beta-actin; promoter; terminator;
 KW resistant; stress; anti-inflammatory drugs.
 XX
 XX Homo sapiens.
 OS
 XX
 PN JP05130819-A.
 XX
 PD 28-MAY-1993.
 XX
 PF 12-NOV-1991; 91JP-0295618.
 XX
 PR 12-NOV-1991; 91JP-0295618.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI; 1993-208254/26.
 DR N-PSDB; AAO43433.

New human ADF transgenic mouse - is resistant to stress and is
 useful for prepn. of anti-inflammatory drugs
 XX
 PS Claim 3; Page 5; 10pp; Japanese.

This sequence represents human ADF. The DNA encoding this sequence
 CC may be used in the production of a transgenic mouse. The transgenic
 CC mouse contains, in the 5' to 3' direction, a human beta-actin promoter,
 CC the human ADF coding gene, a termination codon, a human beta-actin
 CC poly(A) signal and a human beta-actin terminator. The transgenic mouse
 CC is resistant to stress. It is useful in the research of the mechanism
 CC of stress and for the development of anti-inflammatory drugs.

XX


```

CGAGCTGGGCTTGGGTTACAGAGCCGAGACCCCTGAGACTCCGCCACGCTTGGCCAGGCCGCCCGA
470      480      490      500      510      520      530
      30      40      50      60      70      80      90
TTTTTTTGGCTTTTCAANGTCACAGTGTATTTTGGCTTGGAGATCAGCTTACATTATTTCTTGAGTCT
      100      110      120      130      140      150      160
TCCGGGCTCGAGGCTCCAGGGGGGCGGTGGCCGGGGCGCAGCGGGCGAGGTCCGGCGGGGCGC
540      550      560      570      580      590      600
      100      110      120      130      140      150      160
TGCTTCCAAATTTTTCAGATCGGCTCCACAAAACCTCAAAAATCAGGTTGCTCATGAATCCACTTAAC
      170      180      190      200      210      220      230
GCGGCCCCCGCCACAGCCAGCGCGGCGGTGCCAGTTTATTAAGGGAGAGAGCAGCAGCGGCTTGAAGC
610      620      630      640      650      660      670
      170      180      190      200      210      220      230
AGCAAAATTTCTTTTGTGATCTTGAAGATAGG-GTTACCTTCTGGCTTTCTTGAACATCTGAAATGTGGGT
      240      250      260      270      280      290      300
TCTGTTGTGTCCTTGGATCCATTTCCATTCGCTCTTACAGCCGCTCTCAGACTCCAGACCAAGATGTT
680      690      700      710      720      730      740      750
      240      250      260      270      280      290      300
ATTGTTTGATGTGACAAAGTTTCAGCCAGCTTCATGAAAAACAGAAACATCCCTTTGCAAGGACCAACCGT
      310      320      330
GAAGCAGATCGAGACAGGTACGCGCTACCGGGAGAGGCCAGGGGTGCCGC
760      770      780      790      800
      310      320      330
TTCGAAGAAAAATGAACCACTGCGAGTTT

```


SEQ ID NO: 19
Database: A-GeneSeq-032802
AC: NO: AAW95351

Cited
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:20:36 ; Search time 61.07 Seconds

(without alignments)
194.611 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562
Sequence: 1 MVQIKDTNEFKFTLTAAGH.....IFEFGADAKKLEAKTQELM 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A-GeneSeq_032802:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	562	100.0	107	20	AAW95351
2	190	33.8	105	10	AAW92141
3	190	33.8	105	14	AAW37700
4	190	33.8	105	16	AAW23390
5	190	33.8	105	20	AAW25908
6	190	33.8	105	21	AAW30956
7	190	33.8	210	22	AAW05136
8	189	33.6	105	15	AAW47856
9	187	33.3	105	14	AAW42819
10	185	32.9	104	15	AAW47855
11	185	32.9	104	16	AAW23389

12	184	32.7	172	10	AAW4809
13	184	32.7	209	14	AAW31931
14	183	32.6	104	14	AAW42820
15	180	32.0	105	19	AAW69567
16	179	31.9	105	19	AAW69566
17	168	29.9	105	19	AAW69565
18	155	27.6	84	14	AAW42821
19	140	24.9	80	14	AAW42822
20	134	23.8	287	22	AAW61581
21	118	21.0	130	19	AAW70535
22	117.5	20.9	106	22	AAW60229
23	111	19.8	102	20	AAW35279
24	101.5	18.1	107	22	AAW60478
25	100.5	17.9	133	21	AAW10853
26	100.5	17.9	134	21	AAW04653
27	98.5	17.5	178	21	AAW19366
28	95	16.9	144	21	AAW34790
29	94	16.7	122	21	AAW70479
30	93.5	16.6	177	21	AAW16657
31	93.5	16.6	177	21	AAW42782
32	93.5	16.6	275	21	AAW16656
33	93.5	16.6	275	21	AAW42781
34	92.5	16.5	123	22	AAW37102
35	91	16.2	126	21	AAW20450
36	91	16.2	127	21	AAW24028
37	91	16.2	127	21	AAW40198
38	91	16.2	139	21	AAW20449
39	91	16.2	221	21	AAW40197
40	91	16.2	228	21	AAW40196
41	90.5	16.1	115	21	AAW12953
42	90.5	16.1	122	15	AAW65908
43	90.5	16.1	122	21	AAW35810
44	89.5	15.9	127	21	AAW26604
45	89	15.8	109	17	AAW93015

ALIGNMENTS

RESULT 1	AAW95351	standard; Protein; 107 AA.
XX	AAW95351;	
AC	26-APR-1999	(first entry)
XX		
DT	Human adult testis secreted protein fg505_4.	
XX		
DE	Secreted protein; human; testis; fg505_4.	
XX		
KW	Human sapiens.	
XX		
OS	Human sapiens.	
XX		
PN	WO9856909-A2.	
XX		
PD	17-DEC-1998.	
XX		
XX	08-JUN-1998;	98WO-US11822.
PF	05-JUN-1998;	98US-0092722.
XX	PR 11-JUN-1997;	97US-0873218.
PR	(GENY) GENETICS INST INC.	
PA	Agostino MJ, Fechtel K, Howes SH, Jacobs K, Lavallie ER;	
XX	McCoy JM, Racie LA, Spaulding V, Treacy M;	
PI	WPI; 1999-080899/07.	
XX	N-PSDB; AAW9731.	
DR	New polynucleotides encoding secreted human proteins - derived from	
XX	human foetal brain, adult testes, foetal kidney, adult thyroid or	
PT	adult retina cDNA libraries	

XX Claim 31; Page 86; 113pp; English.
PS
XX
CC This is the amino acid sequence of fg505.4, a novel human
CC secreted protein predicted from the nucleotide sequence of a
CC human adult testis cDNA clone (see AAV99731). Database searches
CC indicate some sequence similarity to known sequences. The
CC invention provides cDNA clones (see AAV99721-33) from human adult
CC thyroid, adult retina, adult testis, foetal kidney and foetal
CC brain that encode novel secreted proteins (see AAV95344-53). The
CC polynucleotides and proteins are predicted to have activities
CC which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional,
CC cytokine, cell proliferation/differentiation, immune stimulating
CC (e.g. as vaccines) or immune suppressing, haematopoiesis regulating,
CC tissue growth, activin/inhibin, chemotactic/chemokinetic,
CC haemostatic, thrombolytic, receptor/ligand, antiinflammatory,
CC cadherin/tumour invasion suppressor, and tumour inhibition
CC activities.
CC
XX Sequence 107 AA;
SQ

Query Match 100.0%; Score 562; DB 20; Length 107;
Best Local Similarity 100.0%; Pred. No. 6.2e-64;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVOIIRKDTNEFTFLTAGHKLAVVQSSKRCGPCRMFPVPHLEAETCHITPTFOWF 60
DB 1 mvgilkdneftfltaagkhlaavvqsskrcgpcrkmfpvtheaetchnltptlqmf 60
OY 61 KKSQVTLFSRIKRIICCRSGFMSNLIFPCGADAKKLEAKTOELM 107
DB 61 kksqvtlfsrikriiccrsgfmsnlifpcgadakkleaktqelml 107

RESULT 2

AA92141 ID AAP92141 standard; protein; 105 AA.

AC AAP92141;

DT 25-JUN-1990 (first entry)

DE Recombinant human adult T cell leukaemia derived factor.

KW Human adult T cell leukaemia derived factor; hADF; cancer;

KW immunodeficiency disease.

OS Homo sapiens.

FX Key Location/Qualifiers

FT Region 2..105

FT /note= "Claim 1"

PN EP299206-A.

PD 18-JAN-1989.

PR 10-JUN-1988; 88EP-0109311.

PR 12-JUN-1987; 87JP-0146348.

PA (AJIN) AJINOMOTO KK.

PI Yodoi J, Tagaya Y, Maeda M, Matsui H, Kondo N, Hamuro J;

DR WPI; 1989-016762/03.

DR N-PSDB; AAN93083.

XX Recombinant human adult T cell leukaemia derived factor polypeptide -

PT used for treating cancer, immunodeficiency disease etc.

XX Disclosure; 24pp; English.
PS
XX

CC DNA encoding the polypeptide was sequenced from a gene bank prep. from
CC mRNA isolated from ATL-2 cells from patient with adult T leukaemia
CC virus. Vectors contg. the DNA can be used to transform host cells for
CC prodn. of hADF polypeptide. The polypeptide causes differentiation and
CC induces growth of lymphocytes and fibroblasts. The N-terminal Met is
CC optional. Similar peptides with deletions, sustns. and allelic derivs.
CC may also be used.
CC See also AAP94809.

XX Sequence 105 AA;
SQ

Query Match 33.8%; Score 190; DB 10; Length 105;
Best Local Similarity 36.2%; Pred. No. 2.1e-16;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

OY 1 MVOIIRKDTNEFTFLTAGHKLAVVQSSKRCGPCRMFPVPHLEAETCHITPTFOWF 47
DB 1 mvgilkdneftfltaagkhlaavvqsskrcgpcrkmfpvtheaetchnltptlqmf 60

OY 48 -----TCHITPTFOWFKRSQVTLFSRIKRIICCRSGFMSNLIFPCGADAKKLE 100
DB 61 dcqdvasecevcmbplqfkkqkvq-----eisganekle 98

OY 101 AKTOELM 107

DB 99 atlneliv 105

RESULT 3

AA937700 ID AAR37700 standard; protein; 105 AA.

AC AAR37700;

DT 09-NOV-1993 (first entry)

DE ADF.

KW Human; ADF; transgenic; mouse; beta-actin; promoter; terminator;

KW resistant; stress; anti-inflammatory drugs.

OS Homo sapiens.

PN JP05130819-A.

PD 28-MAY-1993.

PR 12-NOV-1991; 91JP-0295618.

PR 12-NOV-1991; 91JP-0295618.

PA (AJIN) AJINOMOTO KK.

DR WPI; 1993-208254/26.

DR N-PSDB; AAQ43433.

PT New human ADF transgenic mouse - is resistant to stress and is

PS useful for prepn. of antiinflammatory drugs

PS Claim 3; Page 5; 10pp; Japanese.

CC This sequence represents human ADF. The DNA encoding this sequence
CC may be used in the production of a transgenic mouse. The transgenic
CC mouse contains, in the 5' to 3' direction, a human beta-actin promoter,
CC the human ADF coding gene, a termination codon, a human beta-actin
CC poly(A) signal and a human beta-actin terminator. The transgenic mouse
CC is resistant to stress. It is useful in the research of the mechanism
CC of stress and for the development of anti-inflammatory drugs.

SQ Sequence 105 AA;

Query Match 33.8%; Score 190; DB 14; Length 105;
 Best Local Similarity 36.2%; Pred. No. 2.1e-16;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

OY 1 MVOIIKDTEFFKTLTAAGHKLAVVOPSSKRCGPCRMFPVHELA-----47

DB 1 mvqiesktafgealdaagdklvvdatsatwcpckmkpfhlsksynviflevdvd 60

OY 48 -----TCHIKTIPTEFQMEKRSOKVTLFSRIKRIICCYNSGFSNLIFFECGADAKKLE 100

DB 61 dcqdvasecevkcmptfqqfkkqgkv-----efsgankekile 98

OY 101 AKTOELM 107

DB 99 atlneliv 105

RESULT 4

AA72390 ID AAR72390 standard; Protein; 105 AA.

AC AAR72390;

DT 10-NOV-1995 (first entry)

DE Recombinant human ADF.

KW ADF; Inflammation; radiation sickness; fermentation; recombinant.

OS Homo sapiens.

FT Key Location/Qualifiers

PN JP07079780-A.

PD 28-MAR-1995.

PF 20-SEP-1993; 93JP-0233361.

PR 20-SEP-1993; 93JP-0233361.

PA (AJIN) AJINOMOTO KK.

DR WPI; 1995-157850/21.

DR N-PSDB; AA087818.

PT Prepn. of recombinant human ADF by direct expression in E. coli -

PT for treating inflammation and radiation sickness caused by

PT commercially produced free radical(s)

CC The DNA sequence encoding human ADF was inserted into a plasmid which

CC was then used to transform E.coli. The E.coli was then cultured and

CC the recombinantly produced ADF harvested. Human ADF can be used as a

CC treating agent for inflammation and radiation sickness.

XX Sequence 105 AA;

OY Query Match

Best Local Similarity 33.8%; Score 190; DB 16; Length 105;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

OY 1 MVOIIKDTEFFKTLTAAGHKLAVVOPSSKRCGPCRMFPVHELA-----47

DB 1 mvqiesktafgealdaagdklvvdatsatwcpckmkpfhlsksynviflevdvd 60

OY 48 -----TCHIKTIPTEFQMEKRSOKVTLFSRIKRIICCYNSGFSNLIFFECGADAKKLE 100

DB 61 dcqdvasecevkcmptfqqfkkqgkv-----efsgankekile 98

OY 101 AKTOELM 107

DB 99 atlneliv 105

RESULT 5

AA725908 ID AAY25908 standard; Protein; 105 AA.

AC AAY25908;

DT 06-OCT-1999 (first entry)

DE Human thioredoxin protein.

KW Thioredoxin; thioredoxin reductase; human; antisense; primer; metastasis;

KW cytosolic; tumour growth inhibitor; detection; nuclease resistant;

XX phosphorothioate linkage.

OS Homo sapiens.

PN WO9938963-A1.

PD 05-AUG-1999.

PF 29-JAN-1999; 99WO-CA00077.

PR 30-JAN-1998; 98US-0073196.

PA (GENE-) GENESENSE TECHNOLOGIES INC.

PI Lee YS, Wright JA, Young AH;

DR WPI; 1999-469328/39.

DR N-PSDB; AA200544.

PT Antisense oligonucleotides against thioredoxin and thioredoxin

PT reductase genes, useful for inhibiting tumor growth and metastasis

PS Disclosure; Fig 2; 88pp; English.

CC This invention describes novel antisense oligonucleotides against

CC thioredoxin and thioredoxin reductase gene which have cytostatic

CC activity and are useful for inhibiting tumor growth and metastasis

CC in mammals. They may also be used as hybridization probes to detect

CC the presence of the thioredoxin and thioredoxin reductase mRNAs in

CC mammalian cells. They may also be used as molecular weight markers.

CC The antisense oligonucleotides are nuclease resistant due to the

CC presence of phosphorothioate internucleotide linkages. This sequence

XX represents the human thioredoxin protein.

SQ Sequence 105 AA;

Query Match 33.8%; Score 190; DB 20; Length 105;

Best Local Similarity 36.2%; Pred. No. 2.1e-16;
 Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

OY 1 MVOIIKDTEFFKTLTAAGHKLAVVOPSSKRCGPCRMFPVHELA-----47

DB 1 mvqiesktafgealdaagdklvvdatsatwcpckmkpfhlsksynviflevdvd 60

OY 48 -----TCHIKTIPTEFQMEKRSOKVTLFSRIKRIICCYNSGFSNLIFFECGADAKKLE 100

DB 61 dcqdvasecevkcmptfqqfkkqgkv-----efsgankekile 98

OY 101 AKTOELM 107

DB 99 atlneliv 105

```
RESULT 6
AAG03956
ID AAG03956 standard; Protein: 105 AA.
XX
AC AAG03956;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 8037.
XX
KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GSET ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
XX
DR N-PSDB: AAC03962.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 8037; 71pp + CD-ROM; English.
XX
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 105 AA;
XX
XX
Query Match 33.8%; Score 190; DB 21; Length 105;
Best Local Similarity 36.2%; Pred. No. 2.1e-16;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;
```

```
ID AAE05136 standard; Protein: 210 AA.
XX
AC AAE05136;
XX
DT 12-SEP-2001 (first entry)
XX
DE Human T cell-derived colony stimulating factor (TC-CSF).
XX
KW T cell-derived colony stimulating factor; TC-CSF; immunostimulant;
KW acquired immune deficiency syndrome; AIDS; cancer; gene therapy;
KW haematopoietic growth factor; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
XX Peptide 18..28
XX FT /label= signal_peptide
XX FT 29..210
XX FT /label= Mature_human_TC-CSF_protein
XX FT Misc-difference 181
XX FT /note= "Encoded by TAA"
XX FT Misc-difference 188
XX FT /note= "Encoded by TAA"
XX FT Misc-difference 210
XX FT /note= "Encoded by TAA"
XX FT US6254861-B1.
XX
XX PD 03-JUL-2001.
XX
XX PF 24-JAN-1994; 94US-0180371.
XX
XX PR 23-MAY-1989; 89US-0356006.
XX PR 12-JUL-1991; 91US-0729135.
XX PR 19-AUG-1991; 91US-0747784.
XX PR 01-NOV-1991; 91US-0788115.
XX PR 01-JUL-1992; 92US-0906866.
XX
XX PA (CHOUDHURY) CHOUDHURY C.
XX
XX PI Choudhury C;
XX
XX DR WPI: 2001-432041/46.
XX DR N-PSDB: AAD09895.
XX
XX PT New T cell-derived colony stimulating factor for treating
XX immune-compromised patients, including acquired immunodeficiency
XX syndrome patients and certain types of cancer patients -
XX
XX PS Claim 1; Column 49-52; 38pp; English.
XX
XX CC The invention relates to T-cell derived colony stimulating factors
XX (TC-CSF) and nucleic acid molecules encoding them. TC-CSF which
XX is a novel haematopoietic growth factor is useful for treating
XX immune-compromised patients, including acquired immune deficiency
XX syndrome (AIDS) patients and certain types of cancer patients.
XX TC-CSF DNA is used in gene therapy. The present sequence is human
XX TC-CSF.
XX
SQ Sequence 210 AA;
XX
XX
Query Match 33.8%; Score 190; DB 22; Length 210;
Best Local Similarity 36.2%; Pred. No. 5e-16;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;
```

```
RESULT 7
AAE05136
```


XX		Human ADF-polypeptide.	
DE			
XX		ADF: thioredoxin; pancreatitis; amylase; lipase; free radicals;	
KM		protein folding; denaturing; renaturing.	
XX			
OS	Homo sapiens.		
XX			
PN	EP579958-A.		
XX			
PD	26-JAN-1994.		
XX			
PF	18-JUN-1993; 93EP-0109794.		
XX			
PR	19-JUN-1992; 92JP-0161454.		
XX			
PA	(AJIN) AJINOMOTO KK.		
XX			
PI	Asano T, Hamuro J, Hirakawa T, Isono K, Kenmochi T;		
XX			
DR	WPI; 1994-027626/04.		
XX	N-PSDB; AAQ55111.		
PT		Use of a polypeptide having human ADF (thioredoxin) activity -	
PT		for the prophylaxis or therapeutic treatment of pancreatitis or	
PS		diseases which accompany pancreatic disorders	
.XX		Claim 2; Page 9; 15pp; English.	
CC		The ADF polypeptide suppresses the release of the pancreatic enzymes	
CC		amylase and lipase into the blood thereby suppressing pancreatitis.	
CC		The ADF is also capable of eliminating free radicals which can cause	
CC		organ damage, as well as refolding denatured proteins which have	
CC		been denatured by free radicals. The ADF may be isolated from	
CC		cultured human derived cells and prepared by chemical synthesis or	
CC		recombinant techniques	
SQ	Sequence 104 AA:		
Query Match	32.9%; Score 185; DB 15; Length 104;		
Best Local Similarity	35.7%; Pred. No. 8.9e-16;		
Matches 45; Conservative 11; Mismatches 28; Indels 42; Gaps			
OY 2 VOIINDNEFKFLPRAHGKLAVNOFSSKRRCGPCRMPVFHDLAE-----	47		
I : I :	I : I :	I : I :	I : I :
Db 1 VqJieskaftgealaaagkllvvvdfsatwcpckmkfpfhsisekyanvilfedvdgd	60		
I : I :	I : I :	I : I :	I : I :
OY 48 -----TCHIKRTIPFEOMFKRSOKVTLFSRIKRIITICCYNSGMSNLIPEFGADAKRLTA	101		
I : I :	I : I :	I : I :	I : I :
Db 61 cgdvasceevckmpctfgffikkgkyg-----efsganvekilea	98		
I : I :	I : I :	I : I :	I : I :
OY 102 KTOELM 107			
I :	I :		
Db 99 tlnelv 104			
RESULT 11			
AAR72389			
ID AAR72389 standard; Protein; 104 AA.			
XX AAR72389;			
AC			
XX			
DT 10-NOV-1995 (first entry)			
XX			
DE Recombinant human ADF.			
XX			
KM ADF; inflammation; radiation sickness; fermentation; recombinant.			
XX			
OS Homo sapiens.			
XX			
FH Key Location/Qualifiers			
TF Active-site 31..34			

```

XX JP07079780-A.
XX
XX PD 28-MAR-1995.
XX
XX PF 20-SEP-1993; 93JP-0233361.
XX
XX PR 20-SEP-1993; 93JP-0233361.
XX
XX PA (AJIN ) AJINOMOTO KK.
XX
XX DR WPJ: 1995-157850/21.
XX
XX DR N-PSDB; AAQ87817.
XX
XX PT Prepn. of recombinant human ADF by direct expression in E. coli -
XX PT for treating inflammation and radiation sickness caused by
XX PT commercially produced free radical(s)
XX
XX PS Claim 2; Page 7-8; 11pp; Japanese.
XX
XX CC The DNA sequence encoding human ADF was inserted into a plasmid which
XX CC was then used to transform E.coli. The E.coli was then cultured and
XX CC the recombinantly produced ADF harvested. Human ADF can be used as a
XX CC treating agent for inflammation and radiation sickness.
XX
XX SQ Sequence 104 AA;
XX
Query Match 32.9%; Score 185; DB 16; Length 104;
Best Local Similarity 35.7%; Pred. No. 8,9e-16;
Matches 45; Conservative 11; Mismatches 28; Indels 42; Gaps 2;
QY 2 VOIKDINEFTFLAAGHKLAVVQFSSKRCGPCRMPVFHETAE-----47
Db 1 vkgeskrafgedaadgdklvvdfsatwgcgckmkpfhfslekysnvlflevavdd 60
QY 48 -----TCHITPTTFPGMFKSOKVTLEFSRIKRLICCRSGFMSNLIFPCGADAKYLEA 101
Db 61 cgdvassecevcmbptfifkkgqkvg-----efsgankeklea 98
QY 102 KTOELM 107
Db 99 tnelv 104
RESULT 12
AAP94809 AAP94809 standard; protein; 172 AA.
AC AAP94809;
DT 25-JUN-1990 (first entry)
DE Human interleukin 2:human adult T cell leukaemia derived factor fusion
DE protein.
DE XX
KM Human adult T cell leukaemia derived factor; hADF; cancer;
KW Immunodeficiency disease; Interleukin 2; fusion protein.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Region /label= hIL-2
FT Region 71..172
FT FT /label= hADF
PN EP299206-A.
PD 18-JAN-1989.
PE 10-JUN-1988; 88EP-0109311.

```

```

PT haematopoiesis and restoring immune function
XX
PS Disclosure; Page 87-88; 107pp; English.
XX
CC This sequence is encoded by a partial human T cell colony stimulating
CC factor (TC-CSF) clone. Amino acids represented by X in this sequence
CC have been identified as unknown residues. They represent mutations or
CC sequencing errors. TC-CSF polypeptides may be used in pharmaceutical
CC compositions for diagnosis and treatment of immune-compromised
CC patients, including AIDS patients and certain types of cancer patients.
CC TC-CSF and TC-CSF ligands can be used in immunassay. The TC-CSF
CC coding sequence, or portions of it, may be used in hybridisation assay
CC or as primers.
XX
SQ Sequence 209 AA;
XX
Query Match 32.7%; Score 184; DB 14; Length 209;
Best Local Similarity 35.4%; Pred. No. 2.9e-15;
Matches 45; Conservative 11; Mismatches 29; Indels 42; Gaps 2;
QY 1 WQIITKIDNNEFKETFLTAAGHKLAVYQFSSKRCGPKRMFPVFHELA----- 47
Db 76 mvkqskstafgealidaagdklivvdfsatwcpckmkpffhsiseksynvlflevdvd 135
QY 48 -----TCHIKITITFPQMFKSKQKVTLEFSRIKRIIICCRSGFMSNLIPEFGADAKKLE 100
Db 136 dogdvasscevcvcmplqlqffkkgqkvgy-----efsgankekle 173
QY 101 AKTOELM 107
Db 174 atineiv 180
XX
RESULT 14
AAR42820
ID AAR42820 standard; Protein: 104 AA.
XX
AC AAR42820;
XX
DT 04-MAY-1994 (first entry)
XX
DE RECEF 104.
XX
KW Eosinophil cytotoxicity enhancing factor; BCEF; stimulation;
XX diagnosis; therapy; thioredoxin.
XX
OS Homo sapiens.
XX
PN W09320107-A.
XX
PD 14-OCT-1993.
XX
PE 06-APR-1993; 93WO-US03310.
XX
PR 06-APR-1992; 92US-0862832.
XX
PR 06-JUL-1992; 92US-0906842.
XX
PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
PI Balcewicz-sablinska MK, Silberstein DS;
XX
DR WPI; 1993-336837/42.
XX
DR N-PSDB; AAO49403.
XX
XX Shortened Eosinophil Cytotoxicity-Enhancing Factor proteins -
XX having Eosinophil-stimulating activity, for use in diagnosis and
XX therapy
XX
PS Example 3; Page 65; 96pp; English.
XX
CC A 14 kDa molecule designated various names (Eosinophil
CC Cytotoxicity-Enhancing Factor (ECEF); thioredoxin and AdF) has

```

CC been found to support the growth of an EBV transformed B
 CC lymphocyte cell line in the manner of IL-1, to induce the
 CC expression of IL-2 receptors in a large granular lymphocyte cell
 CC line from an ATL patient in the manner of IL-1 and to mediate the
 CC growth inhibitory properties of interferon gamma. Recombinant ECEF
 CC (rECEP) 104 is the full length 104 amino acid polypeptide encoded
 CC by this molecule and was produced for comparison studies against
 CC shortened versions of the polypeptide. Some shortened ECEF's are
 CC useful because they retain eosinophil stimulating activity but lack
 CC proinflammatory diethylol reductase activity. In particular they can
 CC be used for killing tumour cells.

XX Sequence 104 AA;

Query Match 32.6%; Score 183; DB 14; Length 104;
 Best Local Similarity 35.8%; Pred. No. 1,6e-15;
 Matches 44; Conservative 10; Mismatches 27; Indels 42; Gaps 2;

OY 5 IKDNEFKFTLTAAGHKLAIVQSSKRCGCKRMFPVFEHLAE----- 47
 Db 4 Ieskttfgealdaagdklivvdfsatwcpckmnpffhsisekysnvlflevdvdcdg 63
 OY 48 ---TCHIKTPTFQMFKKSKQVTLFSRKRIICCRSGFMSNLIFFECGADAKKLEAKTQ 104
 Db 64 vaseevcmptfqtgfkkgqkvg-----etsgankekleatin 101
 OY 105 ELM 107
 Db 102 elv 104

RESULT 15

AAW69567 ID AAW69567 standard; protein; 105 AA.

XX AC AAW69567;

XX DT 15-OCT-1998 (first entry)

XX DE Human thioredoxin mutant C62S/C69S/C73S.

XX KW Human; thioredoxin; mutant; variant; TRX; AP-1; transcriptional activity;

XX Ref-1.

XX OS Homo sapiens.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT MISC-difference 62 /note= "Cys in wild type"

FT MISC-difference 69 /note= "Cys in wild type"

FT MISC-difference 73 /note= "Cys in wild type"

PN EP853088-A2.

XX PD 15-JUL-1998.

XX PF 14-JAN-1998; 98EP-0100535.

XX PR 14-JAN-1997; 97JP-0004489.

XX PA (ORLY) ORIENTAL YEAST CO LTD.

XX PA (YODO/) YODOI J.

XX PI Yodoi J;

XX DR WPI; 1998-364621/32.

XX New thioredoxin variants with amino acid substitutions for cysteine
 PT except at active centre for increasing stability in non-reducing

PT conditions - used with Ref-1 to enhance AP-1 transcriptional
 PT activity

XX Example 3; Page -; 14pp; English.

CC The present sequence represents a human thioredoxin (TRX) variant
 CC derived from the wild-type sequence (obtained from Genbank), as stated
 CC in the specification. TRX variants can be made stable under non-reducing
 CC conditions by substituting at least one or all Cys residues by other
 CC amino acid residues, except that Cys residues in the active centre of
 CC TRX remain unmodified. Also described in the present invention is a
 CC factor for enhancing the transcriptional activity of AP-1 which
 CC comprises the TRX variant and Ref-1 associated with each other via an
 CC S-S bond. TRX variants can be used with Ref-1 to enhance AP-1
 CC transcriptional activity. TRX variants do not form multimers and are
 CC stable under non-reducing conditions.

XX Sequence 105 AA;

Query Match 32.0%; Score 180; DB 19; Length 105;
 Best Local Similarity 35.4%; Pred. No. 3.9e-15;
 Matches 45; Conservative 11; Mismatches 29; Indels 42; Gaps 2;

OY 1 MVQIKDNEFKFTLTAAGHKLAIVQSSKRCGCKRMFPVFEHLAE----- 47
 Db 1 mvkqiesktatfgealdaagdklivvdfsatwcpckmnpffhsisekysnvlflevdvd 60
 OY 48 -----TCHIKTPTFQMFKKSKQVTLFSRKRIICCRSGFMSNLIFFECGADAKKLE 100
 Db 61 dsqdvaseevkstptfqtgfkkgqkvg-----etsgankekle 98
 OY 101 AKTQELM 107
 Db 99 atinelv 105

Search completed: June 8, 2002, 04:28:04
 Job time: 4048 sec

PD - 8/5/1998

p1 + 2

Mon Jun 24 08:35:20 2002

IntelliGenetics

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file Seq18Inv.res made by tport on Mon 24 Jun 102 8:30:55-PDT.

Query sequence being compared:US-09-746-783-18'	(1-481)
Number of sequences searched:	2
Number of scores above cutoff:	2

Results of the initial comparison of US-09-746-783-18' (1-481) with:
 File : x54539.seq
 File : x77584.seq

Letter	Frequency
N	0
U	0
M	0
B	0
E	0
R	0
O	0
F	0
S	1
E	1
Q	1
U	1
E	1
N	1
C	1

PARAMETERS

	Unary	K-tuple
Similarity matrix	1	4
Mismatch penalty	5.00	Joining penalty
Gap penalty	0.33	Window size
Gap size penalty	1	481
Cutoff score	0	
Randomization group	0	

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	24	25	0.00

Times:	CPU	Total Elapsed
00:00:00	00:00:00	00:00:00

Number of residues:	1384
Number of sequences searched:	2
Number of scores above cutoff:	2

The scores below are sorted by initial score. Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Seq18inv.res

SEE ID NO: 18
INTELLIGENCE, FajDB
AC. NO: X77584, X54539

Page 1

Sequence Name	Description	Length	Init. Opt.	Score	Sig.	Frame
---------------	-------------	--------	------------	-------	------	-------

1. x77584	TOIG of: x77584	check: 4845	583	24	166	0.00	0
2. x54539	TOIG of: x54539	check: 1368	801	24	87	0.00	0

1. US-09-746-783-18' (1-481)
x77584 TOIG of: x77584 check: 4845 from: 1 to: 583

Initial Score	=	24	Optimized Score	=	166	Significance	=	0.00
Residue Identity	=	38%	Matches	=	187	Mismatches	=	294
Gaps	=	10	Conservative Substitutions	=			=	0

CGAGTCTTGAAACATCCTGTTGGTCCTTGGATCATTTCCATTCGGTCTCTTACAGCCGCTCGTCAAGATCCACAA
 10 X 20 30 40 50 60 70

50 60 70 80 90 100 110
AAGTCTAATTTGGCTTGGAGAAC-AGCTTACATTAATCTTGTAGCTTGGCTTCCAAATTTTATAGCATCAGCG
CAGCCAAATGTTGAGACGAGCATCGAGACCAAGACATCTTTTCGAGAACCTTGGACGCTGCAGTGATTAAC
80 90 100 110 120 130 140

120 130 140 150 160 170 180
TCCCAAAAGCTCAAAAGAGTGTGCTCATGAAATCCCATCTGTATAACAGCAATATATCTTTGATTC-
TTT
TTGAGTGGTCACTTCTCAGCCACGCTGTGGGCTGTGCAAAATGATCAAGCCCTTCTCTTTCATPCCCTCTC
150 160 170 180 190 200 210

[illegible]

260 270 280 290 300 310 320 330
CAGCCAGCTCATGAAAAACGAAACATCTTTTGCAGGACACACCGTTTCGAAAGAAATATGAACCACTCT
AGTCAATTCATGCATGCACCAATTCAGTTTTCATTAACAGGACAAAAGGTGGTAAATTTTCGAGGCCAATTA
290 300 310 320 330 340 350 360

CGAGTTGTGTCGCGAGTGTGCAAAANGTTTAAATTCATCTGCTTTTAAATAATGTGACATGA-----
 340 350 360 370 380 390 400
 AGGAAAGCTTGAAACCCACATTATGATGATAGCTATCATGTTTGAAAAATTAACACACCACTTGGCTG
 370 380 390 400 410 420 430

--TTACACGAGGAGTGTGTGAATAATCCCTTGTTGGTTAAGTTGCATCTAGCTGTCTCCTATT
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
ATTTAACCTCTAATTTTTTAATTTTCAAAAATATATAATGACGATAAACCAGTTCATCGGCC

440 450 460 470 480 490 500

480
TTACAGTATCC
| | | |
| | | |
TCACATTAATGCTTACACATTTTAAACCGTCATGTCGTATAGCTTCAA
510 X 520 530 540 550 560

2. US-09-746-783-18' (1-481)
x54539 TOIG of: x54539 check: 1368 from: 1 to: 801

Initial Score	=	24	Optimized Score	=	87	Significance	=	0.00
Residue Identity	=	31%	Matches	=	92	Mismatches	=	196
Gaps	=	1	Conservative Substitutions				=	0

[illegible]

```
CGAGCTGGGCTTGGGTTACAGAGGCCGAAGCCCTGAGCTCCGCCCAAGGCTTGGCCAGCCCGCCCGA
470      480      490      500      510 X      520      530
      30      40      50      60      70      80      90
TTTTTTTTGCTTTTCAAAATGTACACAGTATTTTGGCTTGGAGATCAGCTTACATTAAATCTTGAGTCT
      100      110      120      130      140      150      160
TGCTTCCAAATTTTATGATCGGCTCCACAAAATCAAAAATCAGGTTGCTCATGATCAATCCACTTCTATAC
      170      180      190      200      210      220      230
AGCAAAATTAATCTTTTGAATCTTGAATAAGG-GTTACCTTCTGCTTTTCTTGAACATCTGAATGTGGT
      240      250      260      270      280      290      300
ATTGTTTGAATGTGACAAAGTTTCAGCCAGCTCAATGCAAAACAGGAAACATCTTTTGCAGGGACCAACCGT
      310      320      330
TTGGAAGAAAATTTGAACCACTGCGAGTTT
```

Use for 101

Not Cited
Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:13:55 ; Search time 1874.6 seconds
(without alignments)
5369.300 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481
Sequence: 1 GCATACCTGTAATAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

GenBml: 1: gb_ba: 2: gb_hlg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_sts: 28: em_un: 29: em_vl: 30: em_hlg_hum: 31: em_hlg_inv: 32: em_hlg_other: 33: em_hlg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	ID	Description
1	114	194835	9	AL158158	Human DNA
2	97	20.2-214852	9	CNS05Tcz	Human chr
3	79.4	16.5	255	AF055241	Homo sapi
4	63.8	13.3	350	AF104105	Bos tauru
5	62.2	12.9	532	OATHIORD	O.aries mRN
6	61.6	12.8	382	AX321985	Sequence
7	61.4	12.6	371	AF382821	Sus scrofa
8	60.6	12.6	583	E01915	CDNA encodi
9	60.6	12.6	583	HSATLRED	Homo sapi
10	60	12.5	529	BC003377	Homo sapi
11	59.8	12.5	630	AR160057	Sequence
12	59.6	12.4	371	AF313911	Homo sapi
13	58.8	12.4	477	HUMY146D09	Homo sapi
14	58.8	12.2	439	AY004872	Homo sapi
15	58.4	12.1	501	HUMYTHD	Human thior
16	57.8	12.0	364	AF353204	Callichr
17	57.8	12.0	591	AF321769	Ophiophag
18	57.2	11.9	421	AF276919	Homo sapi
19	56.8	11.8	506	MACYRDA	Macaca mula
20	56.8	11.8	630	AR160060	Sequence
21	56	11.6	335	E05024	Human ADF
22	56	11.6	315	E05929	DNA encodin
23	56	11.6	315	E08965	DNA encodin
24	55.2	11.5	562	AF146023	Homo sapi
25	55.2	11.5	661	CHKTHD	Chicken th
26	55.2	11.5	117930	AC073271	Homo sapi
27	55.2	11.5	161081	AC022897	Homo sapi
28	55.2	11.5	163319	AC108512	Homo sapi
29	55	11.4	1764	HSAA20593	Homo sapi
30	53.6	11.1	535	AB022431	Equus cab
31	53	11.0	312	E06575	CDNA encodi
32	53	11.0	312	E08964	DNA encodin
33	51.2	10.6	1210	AF146024	Homo sapi
34	51.2	10.6	17750	AC005874	Human c11b
35	51.2	10.6	177864	AF134471	Homo sapi
36	51.2	10.6	178964	AC012468	Homo sapi
37	51	10.6	4938	AY069493	Drosophil
38	50.2	10.4	534	BC010756	Mus muscu
39	50.2	10.4	538	MMTRX	M. musculus
40	50	10.4	362	A84204	Sequence 29
41	50	10.4	349980	AX344558	Sequence
42	49.8	10.4	311	AF159977	Buthus ma
43	49.8	10.4	1577	BC011971	Homo sapi
44	49.8	10.4	63374	AC102559	Mus muscu
45	49.4	10.3	5383	AB027537	Xenopus 1

ALIGNMENTS

RESULT 1
AL158158/c
LOCUS
DEFINITION Human DNA sequence from clone Rpl1-427/L11 on chromosome 9q31.2-32, complete sequence.
ACCESSION AL158158
VERSION AL158158.14 GI:13559997
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 194835)
AUTHORS Williams, S.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 6, 2001 this sequence version replaced gi:12733508.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

COMMENT

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>

RP11-427L11 is from the library RP11-11.2 constructed by the group of Pletier de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-427L11.

FEATURES

source

Location/Qualifiers

1..194835

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/map="q31.2-32"

/clone="RP11-427L11"

/clone_id="RP11-11.2"

1..224

/note="L1P repeat: matches 3..226 of consensus"

326..615

/note="Alusg repeat: matches 1..299 of consensus"

627..772

/note="Alusg/x repeat: matches 158..301 of consensus"

929..1168

/note="L1M4 repeat: matches 5149..5403 of consensus"

1195..1519

/note="Alusx repeat: matches 1..311 of consensus"

1698..1726

/note="MER3A repeat: matches 16..44 of consensus"

1728..1868

/note="Alusg/x repeat: matches 150..302 of consensus"

1887..1932

/note="MER3A repeat: matches 61..106 of consensus"

1966..2242

/note="L1M3A repeat: matches 5614..5894 of consensus"

complement(1992..2500)

/note="match: GSS: Em:AQ172930"

2308..2619

/note="Alusg repeat: matches 1..313 of consensus"

2707..2822

/note="L1M3A repeat: matches 6017..6137 of consensus"

3336..3403

/note="34 copies 2 mer cc 66% conserved"

3420..3725

/note="Alusx repeat: matches 1..300 of consensus"

4319..4467

/note="MER3B repeat: matches 4..175 of consensus"

4884..5309

/note="L2 repeat: matches 166..616 of consensus"

5596..5870

/note="L1M9 repeat: matches 6027..6305 of consensus"

5871..5938

/note="Alus/FLAM repeat: matches 1..68 of consensus"

5967..6648

/note="L2 repeat: matches 1284..2090 of consensus"

6649..6948

/note="Alusg repeat: matches 1..300 of consensus"

6949..7003

/note="L2 repeat: matches 2090..2136 of consensus"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

7072..7482

/note="L2 repeat: matches 2295..2695 of consensus"

7588..7789

/note="MIR repeat: matches 12..226 of consensus"

complement(8285..8469)

/note="match: STS: Em:G05863"

complement(8329..8495)

/note="match: GSS: Em:B65750"

complement(8329..8496)

/note="match: GSS: Em:A0227987"

complement(8339..8502)

/note="match: GSS: Em:A0110861"

8354..8529

/note="match: STS: Em:G54088"

join(8370..8532,9091..9156,15133..15192,15671..15768)

/note="match: STS: Em:G29920"

8591..8662

/note="36 copies 2 mer ta 81% conserved"

9732..9893

/note="Alud repeat: matches 130..287 of consensus"

complement(9737..9894)

/note="match: GSS: Em:A0281101"

complement(9886..10119)

/note="match: GSS: Em:B53125"

complement(9897..10091)

/note="match: STS: Em:MSA108WE1"

complement(9907..10123)

/note="match: GSS: Em:B17453"

9932..9967

/note="18 copies 2 mer ta 100% conserved"

9973..10105

/note="Alud repeat: matches 1..133 of consensus"

10133..10392

/note="Alusx repeat: matches 38..299 of consensus"

10183..10319

/note="match: GSS: Em:AQ782105"

10404..10492

/note="MER3B repeat: matches 7..97 of consensus"

10992..11078

/note="match: GSS: Em:AQ544271"

11354..11653

/note="Alusx repeat: matches 1..300 of consensus"

11742..12929

/note="Tiger3b repeat: matches 2..1231 of consensus"

12987..13146

/note="MIR repeat: matches 89..248 of consensus"

13153..13468

/note="Alusx repeat: matches 1..312 of consensus"

13678..13848

/note="L2 repeat: matches 2537..2703 of consensus"

14394..14609

/note="MIR repeat: matches 12..257 of consensus"

15816..15902

/note="match: GSS: Em:B46502"

13900..15947

/note="24 copies 2 mer aa 79% conserved"

16257..16558

/note="Alusg repeat: matches 1..302 of consensus"

16956..17220

/note="Alusg repeat: matches 38..304 of consensus"

17415..17488

/note="37 copies 2 mer aa 71% conserved"

18552..18845

/note="Alusg repeat: matches 1..293 of consensus"

19629..19689

/note="L2 repeat: matches 2685..2743 of consensus"

19779..21330

/note="Cpg island"

/evidence=not_experimental

21481..21679

/note="MIR repeat: matches 48..262 of consensus"

22591..22818

/note="MIR repeat: matches 23..258 of consensus"

```
misc_feature complement(23067..23458)
              /note="match: GSS: Em:B43385"
misc_feature complement(23074..23458)
              /note="match: GSS: Em:B43132"
repeat_region repeat_region
              /note="MER5A repeat: matches 17..180 of consensus"
misc_feature 24008..24365
              /note="match: GSS: Em:AQ088039"
repeat_region repeat_region
              /note="MIR repeat: matches 105..202 of consensus"
repeat_region repeat_region
              /note="LIMB8 repeat: matches 6145..6287 of consensus"
repeat_region repeat_region
              /note="MER50 repeat: matches 1..436 of consensus"
repeat_region repeat_region
              /note="MER50 repeat: matches 640..734 of consensus"
repeat_region repeat_region
              /note="LIME1 repeat: matches 5212..5366 of consensus"
repeat_region repeat_region
              /note="ALUSP/q repeat: matches 174..260 of consensus"
repeat_region repeat_region
              /note="LIME1 repeat: matches 5347..6036 of consensus"
repeat_region repeat_region
              /note="LIMB6 repeat: matches 4066..4176 of consensus"
repeat_region repeat_region
              /note="ALUSX repeat: matches 1..310 of consensus"
repeat_region repeat_region
              /note="LIMB6 repeat: matches 4176..4455 of consensus"
repeat_region repeat_region
              /note="ALUB repeat: matches 3..132 of consensus"
repeat_region repeat_region
              /note="ALUS30 repeat: matches 1..297 of consensus"
repeat_region repeat_region
              /note="ALUSX repeat: matches 1..297 of consensus"
              28531..28677
```

```
Query Match 23.7%; Score 114; DB 9; Length 194835;
Best Local Similarity 92.3%; Pred. No. 5.8e-16;
Matches 120; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 330 TTCATGAGCAACGATTTTGAAGTTTGTGAGCCGATGCTAAATAATGGAAGCCAAAG 389
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67911 TGCATTGTGACAGATTTTGAAGTTTGTGAGCCGATGCTAAATAATGGAAGCCAAAG 67852

QY 390 ACTCAGAAATTAATGATGATGATCTCCAAAGCAAAATACACTTGATGATTTGAAAAAG 449
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67851 ACTCAGAAATTAATGATGATGATCTCCAAAGCAAAATACACTTGATGATTTGAAAAAG 67792

QY 450 CAAAAAATAA 459
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67791 CAAAGACAAA 67782
```

```
RESULT 2
CNS05TC2 214852 bp DNA linear PRI 29-MAY-2001
LOCUS Human chromosome 14 DNA sequence BAC R-322L17 of library RCI-11
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL355888
VERSION AL355888.3 GI:14272176
KEYWORDS HMG; HTGS_ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 214852)
AUTHORS Hellig, R., Pett, J. L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Cattelino, L., Barbe, V., Pelletier, E., Aftigenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Crnaud, C.,
Gyapay, G., Saurin, W. and Weissenbach, J.
TITLE Unpublished
JOURNAL Sequencing of the human chromosome 14
REFERENCE 2 (bases 1 to 214852)
AUTHORS Genoscope.
TITLE Direct Submission
```

JOURNAL Submitted (29-MAY-2001) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On May 31, 2001 this sequence version replaced gi:9886699.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SegRef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-305B6
Downstream BAC (overlapping the SP6 end) : R-99L13 (AC-AL161752)
----- Summary Statistics
Assembly program: Phrap; Statistics
Quality coverage: 6.02x in Q20 bases; sum-of-contigs

Overall quality chart :
Range : bases

0	1	2
10	19	29
20	29	127
30	39	494
40	49	4696
50	59	14835
60	69	19064
70	79	37628
80	89	65006
90	99	72971

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES
source Location/Qualifiers

1..214852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-322L17"
/clone_1b="RPCR-11"
181957..182185
/note="matching EMBL:G02797
RHdb:RH53844
RHdb:RH3727
dbSTS:STS14441
Identified using the e-PCR software (G. Schuler)"
BASE COUNT 74248 a 38130 c 37571 g 64903 t
ORIGIN

Query Match 20.2%; Score 97; DB 9; Length 214852;
Best Local Similarity 87.6%; Pred. No. 3.7e-12;
Matches 106; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
QY 339 AACCTGATTTTGAAGTTTGTGAGCCGATGCTAAATAATGGAAGCCAAAGCTCAAGAA 398
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91231 AACGACACATTTAGATTGTTGGGGCGATGCTAAATAATGGAAGCCAAAGATTCAGAA 91290

QY 399 TTAATGTAAGCTGATCTCCAAAGCAAAATACACTTGACATTTGAAAAAGCAAAAAA 458
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 91291 TTAATGCAAGCTGATCTCCAAAGCAAAATACACTTGACATTTGAAAAAGCAAAAAA 91350

QY 459 A 459
    |
Db 91351 A 91351
```

```
RESULT 3
AF065241 AF065241 255 bp mRNA linear PRI 27-MAY-1998
LOCUS Homo sapiens thioresdoxin delta 3 (TXN delta 3) mRNA, partial cds.
DEFINITION AF065241
ACCESSION AF065241.1 GI:3153858
VERSION
```


Query Match 12.8%; Score 61.4; DB 4; Length 371;
Best Local Similarity 52.8%; Pred. No. 0.00054;
Matches 181; Conservative 0; Mismatches 156; Indels 6; Gaps 2;

```

OY 75 GGTGAATCAGGTACAGATTATTAAGACAGCAATGAATTTAAACATTTTGGACAGCT 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 GCTGCCAAGATGTGATGACAGATGACAGCAAGTACGCTTTTCAGAGAGCCCTTGAAACGT 61
OY 135 GCCGACACAAACATGCGAGTGGTTCATTTTCTTGGAAACGGTGGTCCCTGCAGAAAG 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 GCAGGAGGAAACCTGTGTGTGTCATTTTCAGCCACGTGGTGGGCTTGCAAAATG 121
OY 195 ATGTTCCCTGTTTTCATGAGCTGCTGAACCTTCTCAGATCAATCAAAACATACCCATTTT 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 ATCAAGCCTTTCTTCCATTTCTCTCTGAAAAGTATTCAGATGCTGCTTCTTAAGTA 181
OY 235 CAGATGTTCAAGAAAAGCCAGAGTAAACCTATTCTCAAGATCAAAAGATTAATTTGC 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 GATGTG---GATGACTGTGAGATGTGCTTCAGAGTGTGAAGTCAAAATGATGCCAAC 238
OY 315 TGTATAGAGTGAATGATAGCAACCTGATTTTGTGATTTTGGAGCCGATCTTAA 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 TTCCAGTTTAA---AAGGACAGAAAGGTGGGTGAATTTTCTGAGCTAACAGGAA 295
OY 375 AAGTTGAAGCCAGACTCAAGATTAATTAATGAAGCTGATCTCC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 AACTTGAAGCCACCATTAATGAATTAATCATATGTTTC 338

RESULT 8
LOCUS E01915 583 bp RNA linear PAT 29-SEP-1997
DEFINITION cDNA encoding human adult T-cell leukemia derived factor.
ACCESSION E01915
VERSION E01915.1 GI:2170164
KEYWORDS JP 1989085097-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 583)
AUTHORS Yodoi,J., Tagaya,A., Maeda,M., Matsui,Y., Kondo,N. and Hamuro,J.
RECOMBINANT HUMAN ADF
Patent: JP 1989085097-A 1 30-MAR-1989;
AJINOMOTO CO INC, YODOI JIYUNJI
OS Homo sapiens (man)
PN JP 1989085097-A/1
PD 30-MAR-1989
PE 31-MAY-1988 JP 1988134218
PI 12-JUN-1987 JP 87P 146348
PI YODOI JIYUNJI, TAGAYA ATSUSHI, MAEDA MICHIOKI, MATSUI YUTAKA,
PI KONDO NOBUO, HAMURO JUNJI
PC C12P21/02.C07K13/00.C12N1/16.C12N1/20.C12N5/00, PC
C12N15/00/A61K37/02,
PC A61K37/02.(C12P21/02.C12R1.19).(C12P21/02.C12R1.865), PC
(C12P21/02.C12R1.91);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue-type=Leukemia;
CC *source: cell_line=ATL-2;
CC *source: clone=+20;
FH Key Location/Qualifiers
FH 5'UTR <1..80
FH mat_peptide 81..395
FH /product='Human adult T-cell leukemia derived
FH factor'
FH /note='ADF'
FH CDS 81..398
FH /product='Human adult T-cell leukemia derived
FH factor'
FH /note='ADF'
FH 3'UTR 399..>583.
FH Location/Qualifiers
FEATURES

```

```

source 1..583
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 178 a 111 c 120 g 174 t
ORIGIN
Query Match 12.6%; Score 60.6; DB 6; Length 583;
Best Local Similarity 52.1%; Pred. No. 0.00079;
Matches 185; Conservative 0; Mismatches 164; Indels 6; Gaps 2;

OY 62 TCAGACCTCCCTGGTGAATCATGCTACAGATTATTAAGACAGATGAATTTAAAC 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 TCGTCAGACTCCAGACGCAAGATGTCAGACAGATTCAGAGCAAGACTGCTTTCAGGA 118
OY 122 ATTTTTCAGACGTCGCGACACAAACCTGCAGTGTTCATTTTCTTGGAAACGGTGG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 AGCCTTGACGCTGAGAGTGAATTAACCTGTAGTACTTCTCAGCAGCTGCTGG 178
OY 182 TCCTGCAAAAGATGTTTCTGTTTCCATGAGCTGCTGAACCTGTCAATCAAAAC 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 GCCTTGCAAAATGATCAAGCCTTTCTTCAATCCCTCTGAAAAGTATTCACAGTAT 238
OY 242 AATNCCCATTTTCAGATGTTCAAGAAAAGCCAGAAAGTAACCCATTCACAGATCAA 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 ATTCCTGAAGTAGATGTG---GATGACTGTGAGATGTCTTCAGAGTGAAGTCAA 295
OY 302 AAGATTAATTTGCTGTTTATAGAGTGAATTCATGACCAACCTGATTTTGGATTTTGG 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 ATGCATGCCACATTTCCAGCTTTTAAAG---AAGGACAAAAGGTGGTGAATTTTCGG 352
OY 362 AGCCGATGCTAATAAATTGGAAGCCAGACTCAAGATTAATTAATGAAGCTGATCTC 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 AGCCAAATGAAGAAAAGCTTGAAGCCACCATTAATGAATTAATCATATGTTTC 407

RESULT 9
LOCUS HSATLRED 583 bp mRNA linear PRI 09-FEB-1994
DEFINITION H.sapiens mRNA for ATL-derived factor/thioredoxin.
ACCESSION X77584
VERSION X77584.1 GI:453963
KEYWORDS thioredoxin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 583)
AUTHORS Tagaya,Y., Maeda,Y., Mitsui,A., Kondo,N., Matsui,H., Hamuro,J.,
Brown,N., Arai,K.I., Yokota,T., Wakasugi,H. and Yodoi,J.
ATL-derived factor (ADF), an IL-2 receptor/Tac inducer homologous
to thioredoxin; possible involvement of dithiol-reduction in the
IL-2 receptor induction
Erratum: [[published erratum appears in EMBO J 1994 May
1;13(9):2244]]
2 (bases 1 to 583)
REMARK Tagaya,Y.
JOURNAL Direct Submission
EMBO J. 8 (3), 757-764 (1989)
89251607
MEDLINE
REMARK
JOURNAL Erratum: [[published erratum appears in EMBO J 1994 May
1;13(9):2244]]
2 (bases 1 to 583)
REFERENCE Tagaya,Y.
AUTHORS Tagaya,Y.
TITLE Submitted (07-FEB-1994) Y. Tagaya, Division of Allergy Immunology,
Dept of Medicine, Cornell Medical Center, 525 East 68th Street, Km.
1C 907, New York, NY 10021, USA
JOURNML Location/Qualifiers
FEATURES
source 1..583
/organism='Homo sapiens'
/db_xref='taxon:9606'
/chromosome='1'
/cell_line='ATL-2 (human HTLV-I transformed T cell line)'
/clone_idb='ATL-2/Lgt11'
81..398
/codon_start=1
/product='ATL-derived factor/thioredoxin'

```


Probable match to Homo sapiens protein P10599 (PID:g2982059) Human
Thioredoxin (Oxidized With Diamide) Oxidoreductase, Dimer.
Thioredoxin, X-Ray Crystallography, Electron Transport Mol_id: 1;
Molecule: Thioredoxin; Chain: Null

The location of this clone is unknown.
Location/Qualifiers

FEATURES
source
1..477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="142289"
/clone_lib="Soares_placenta_1NHP"
misc_feature
13..327
/note="Probable match to Homo sapiens protein J04026 (PID:g339649)"
misc_feature
13..327
/note="Probable match to Homo sapiens protein P10599 (PID:g2982059)"
BASE COUNT 159 a 84 c 96 g 138 t
ORIGIN

Query Match 12.4%; Score 59.6; DB 9; Length 477;
Best Local Similarity 52.3%; Pred. No. 0.0013;
Matches 181; Conservative 0; Mismatches 159; Indels 6; Gaps 2;

QY 72 CCTGCTATCATGTTAAGATTTATTAAGACACGATGATTTAAACATTTTGGACA 131
DB 1 CCAGACGCCAAGATGGTGAAGCAGATCGAGACAGCTGCTTTCAGAAAGCCTTGAC 60
QY 132 GCTGCCGACACAACTCCAGATGTTCAATTTTCTCGAAACGGTGGTCCCTGCAA 191
DB 61 GCTGCGAGTGAATTAATTTGATGATTTCTCAGCCAGCGTGCTGGCCCTTGCAA 120
QY 192 AGAGTGTTCCTGTTTCCATGAGCTGCTGAAACTTTCACATCAAAACATATCCACA 251
DB 121 ATGATCAAGCCCTTCTTCTTCCCTCTCGAAATATTCACAGTGATATTCCTGAA 180
QY 252 TTTCAGATTTCAAGAAAGCCAGAGTAACTTCTCAGATTCAAATCAAAATTAAT 311
DB 181 GTAGATGTC---GATGACTGTGAGATGTTGCTTCAGAGTGAAGTCAATGCAATCCA 237
QY 312 TCGTGTATAGAGTGTATTCAGACCAACCTGATTTTGTGAGTGGAGCCGATGCT 371
DB 238 ACATTCACATTTTAAAG---AAGGACAAAGAGTGGGTAATTTCTGAGCCATTAAG 294
QY 372 AAAAAATGGAGCCAGACTCAAGATTAATTAAGCTGATCTCC 417
DB 295 GAAAGCTTGAAGCCACATTAATGATTAATCAATGTTTTC 340

RESULT 14

LOCUS AT004872 439 bp mRNA linear PRI 21-JUL-2000
DEFINITION Homo sapiens thioredoxin (TXN) mRNA, complete cds.
ACCESSION AY004872
VERSION AY004872.1 GI:9508996

KEYWORDS
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 439)
AUTHORS Liu, A. and Lou, M.F.
TITLE Cloning, purification and characterization of human lens
thioredoxin

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 439)
AUTHORS Liu, A. and Lou, M.F.

TITLE Direct Submission
JOURNAL Submitted (14-JUL-2000) Veterinary & Biomedical Sciences,
University of Nebraska at Lincoln, 120 VBS, East Campus, Lincoln,
NE 68583-0905, USA

FEATURES
source
Location/Qualifiers
1..439
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HLE-B3"
/cell_type="epithelial"
/tissue_type="lens"
1..439
/gene="TXN"
CDS
11..328
/gene="TXN"
/note="oxidoreductase"
/codon_start=1
/product="thioredoxin"
/protein_id="AAF87085.1"
/db_xref="GI:9508997"
/translation="MWKQIESKTAPEQALDAAGDKLVVDFSATWCGPKMKPEFHS
LSEKYSNVIFLEVDVDDCODVASSECEVKCMPTFQFFKKGQKVEFGANKRELEATIN
ELV"

BASE COUNT 138 a 78 c 95 g 128 t
ORIGIN

Query Match 12.2%; Score 58.8; DB 9; Length 439;
Best Local Similarity 52.7%; Pred. No. 0.002;
Matches 176; Conservative 0; Mismatches 152; Indels 6; Gaps 2;

QY 84 ATGTCACAGATTTTAAAGACAGCAATGAAATTTAAACATTTTGGACGCTGCCGACAC 143
DB 11 ATGTCAGACAGATTCGAGACAGCTGCTTTCAGAAAGCCTTGACAGCTGAT 70
QY 144 AAATCGACATGGTCAATTTCTCGAAACGGTGGTCCCTGCAAAAGATGTTCT 203
DB 71 AAACCTGTACTAGTGTGCTTCTCAGCCAGCTGGTGGCTTGCAAAATGATCAACCT 130
QY 204 GTTTCATGAGCTGGCTGGAACCTTGCACATCAAAACATATTCAGATGTC 263
DB 131 TTCTTCATTCCTCTCTGAAAGTATTCACAGTGATATTCCTGAAATGATGTC--- 187
QY 264 AAAAAAGCCAGAGGTAACCTATTTCTCAAGATCAAAAGATTAATTTCTGTTAAGA 323
DB 188 GATGACTGTGAGATGTGCTTCAGAGTGAAGTCAAAATGCAATTCACATTCACATTT 247
QY 324 AGTGATTCATGAGCAACCTGATTTTGTGAGTGGAGCCGCTGAATAAATGGA 383
DB 248 TTTAAG---AAGGACAAAGAGTGGTGAATTTTCTGAGCCATTAAGAAAGCTTGA 304
QY 384 GCCAAGACTCAAGATTAATTAAGCTGATCTCC 417
DB 305 GCCACCATTAATGATTAATGATTAATGATTTTC 338

RESULT 15

LOCUS HUMTHD 501 bp mRNA linear PRI 23-AUG-1995
DEFINITION Human thioredoxin (TXN) mRNA, complete cds.
ACCESSION J04026
VERSION J04026.1 GI:339648

KEYWORDS
SOURCE Human cDNA to mRNA.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 501)
AUTHORS Wolman, E.E., d'Aurion, L., Rimsky, L., Shaw, A., Jacquot, J.P.,
Wingfield, P., Graber, P., Desarsy, F., Robin, P., Galibert, F.,
Bertoglio, J., and Fradeliz, D.

TITLE Cloning and expression of a cDNA for human thioredoxin

JOURNAL J. Biol. Chem. 263 (30), 15506-15512 (1988)
REFERENCE 2 (bases 1 to 501)
AUTHORS Wolman, E.E.

TITLE Direct Submission

JOURNAL

Submitted (24-AUG-1988) E.E. Wollman, Centre National de la
Recherche Scientifique, UA 1156/Institut National de la Sante et de
la Recherche Medicale/Institut Gustave Roussy, Villejuif, France

FEATURES

Location/Qualifiers

1..501

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="3p12-p11"

/cell_line="3B6"

/tissue_type="EBV transformed lymphocytes"

1..501

/gene="TXN"

64..381

/gene="TXN"

/codon_start=1

/product="thioredoxin"

/protein_id="AA74596.1"

/db_xref="GI:339649"

/db_xref="GDB:G00-120-475"

/translation="MKQIESKTAFOEALDPAQDKLVVDEFSATWGPCKINPFHS
LSEKYSNVIFLEVVDVDCQDVASECEVKCTPEQFQFKGQKVEFSGANKEKLEATIN
ELY"

gene

CDS

BASE COUNT

149 a 100 c 105 g 147 t
ORIGIN 1 bp upstream of EcoRI site.

Query Match

12.1%; Score 58.4; DB 9; length 501;

Best local similarity 51.7%; Pred. No. 0.0025;

Matches 184; Conservative 0; Mismatches 166; Indels 6; Gaps 2;

```
QY 62 TCAGACTTCCTGGTGAATCATGCTGACAGATTATTAAGACACGATGAATTTAAAC 121
    || || || || || || || || || || || || || || || || || || || ||
Db 42 TCGTAGACTCCAGCAGCGCAAGATGTGTAAGCAGATCGAGCAAGACTGCTTTCAGGA 101

QY 122 ATTTTTCAGAGCTCCGCGACACAAAATCGACAGTGTCAATTTCTTGAACGGGTGG 181
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 102 AGCCTTGAGCGCTGAGGTGATTAACCTGTAGTACTTCTCAGCAGCTGCTGTGG 161

QY 182 TCCCTGCAAAAGATGTTTCTGTTCATGAGCTGCTGAACCTGTCAATCAAAAC 241
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 162 GCCTTGCAAAATGATCAACCTTCTTTCATTCCTCTGTAAGATTCCTCAAGTGAAT 221

QY 242 AATACCCACATTTAGATGTCAGAAAGCAGAGTAACCTATTCACAGATCA 301
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 222 ATTCCTGAGATGATGAGTGG--GATGACTGTGAGATGTGCTTCAGAGTGAAGTCAA 278

QY 302 AAGATAATTTGCTGTATTAGAGTGAATGATGAGCAACTGATTTTGAATTTGGTGG 361
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 279 ATGACAGCGCACATTCACGATTTTAAAG--AAGGACAAAAGGTGGGTAATTTCTGG 335

QY 362 AGCGGATGCTAAATAATTTGAGCCAGACTCAAGATTAATGTAAGCTGATCTCC 417
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 336 AGCCAATTAAGGAAAAGCTTGAAGCCACCATTAATGAATTAATGATCATGTTTTC 391
```

Search completed: June 8, 2002, 03:20:32
Job time: 3997 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:15:20 ; Search time 78.88 seconds
(Without alignments)
1497.841 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481
Sequence: 1 GCATACCTGTAATTAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
4: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
5: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
6: /cgn2_6/prodata/2/lna/5B_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.4	14.6	631	4	US-09-385-982-174
2	60	12.5	624	4	US-09-385-982-526
3	60	12.5	630	4	US-08-180-371-5
4	56.8	11.8	630	5	PCR-US92-05707-5
5	54.2	11.3	605	4	US-08-180-371-17
6	47	9.8	5173	1	US-09-385-982-485
7	46.4	9.6	1461	3	US-08-242-677-1
8	46.4	9.6	1461	3	US-08-722-1264-4
9	46.4	9.6	1461	5	PCR-US95-04258-4
10	45.6	9.5	7218	1	US-08-232-463-14
11	45.6	9.5	1604	3	US-08-665-966-9
12	45.6	9.5	1604	3	US-09-041-780-9
13	45.4	9.4	2209	1	US-08-514-014-1
14	45.4	9.4	2209	2	US-08-833-823-1
15	45.4	9.4	2280	3	US-08-813-150-1
16	45.2	9.4	1910	4	US-08-974-691-1
17	44.4	9.2	270	2	US-08-520-678A-30
18	44.4	9.2	270	4	US-08-897-126-30
19	44.4	9.2	5555	1	US-08-484-438-3
20	44.4	9.2	19557	5	PCR-US92-06300-1
21	44.2	9.2	1046	1	US-08-361-467B-4
22	44.2	9.2	1046	1	US-08-484-332C-4
23	44.2	9.2	1117	4	US-09-247-373B-33
24	43.8	9.1	2158	1	US-07-602-608-1
25	43.8	9.1	2158	1	US-08-261-578-1
26	43.6	9.1	2674	4	US-09-817-180-1
27	43.4	9.0	960	3	US-09-248-335-57

28	43.4	9.0	1100	4	US-07-861-458C-4	Sequence 4, Appl
29	43.4	9.0	2625	4	US-09-245-041-18	Sequence 18, Appl
30	43	8.9	1577	4	US-08-821-994-59	Sequence 59, Appl
31	42.8	8.9	577	4	US-09-385-982-203	Sequence 203, Appl
32	42.6	8.9	1129	4	US-09-227-357-40	Sequence 40, Appl
33	42.6	8.9	1700	2	US-08-897-340-4	Sequence 4, Appl
34	42.6	8.9	1700	3	US-09-252-329-4	Sequence 11, Appl
35	42.4	8.8	1798	4	US-09-797-906-1	Sequence 4, Appl
36	42.4	8.8	2114	1	US-07-803-622E-6	Sequence 6, Appl
37	42.4	8.8	2628	1	US-08-143-219-1	Sequence 1, Appl
38	42.2	8.8	1493	1	US-08-340-820-24	Sequence 24, Appl
39	42.2	8.8	1493	1	US-08-593-535-24	Sequence 24, Appl
40	42.2	8.8	1921	2	US-08-557-128-11	Sequence 11, Appl
41	42.2	8.8	2550	6	5258287-23	Patent No. 5258287
42	42.2	8.8	2837	2	US-08-993-228-11	Sequence 11, Appl
43	42	8.7	880	4	US-09-247-373B-39	Sequence 39, Appl
44	42	8.7	1420	2	US-08-909-965C-3	Sequence 3, Appl
45	42	8.7	12980	3	US-08-811-566-5	Sequence 5, Appl

ALIGNMENTS

```
RESULT 1
US-09-385-982-174
; Sequence 174, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 344
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 174
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(631)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-174
```

Query Match 14.6%; Score 70.4; DB 4; Length 631;
Best Local Similarity 53.7%; Pred. No. 1.1e-08;
Matches 191; Conservative 0; Mismatches 161; Indels 4; Gaps 2;

QY	62	TCACACTTCCTCCGTGTAATCATGATGATTTTAAGACACGATGATTTAAAC	121
DB	116	tcgacagactccagcagcaagatgctgagcagatcagcagcagctgttcgcgga	175
QY	122	ATTTTACAGCTCCCGGACACAACTCGAGTGGTTCATTTTCGAAACGGTGG	181
DB	176	agctctgagcctcagctgataactctgctgactctcagcagcagctgtgtg	235
QY	182	TCCCTGCAAAAGATGTTTCTGTTTCCATGAGCTGGCTGAACCTTGTCATCAATCAAC	241
DB	236	gccttgaataatgatacagccttcttcacccctctctgaaagatctcaacgctg	295
QY	242	AATGCCACATTTTGAATGTTTCAGAAAAGCACAAGTACCTTATTTCAAAATCAA	301
DB	296	attccttgaagt-aggatgtgactgtcagatgtgtgtcctcaaaagtgtgaagtc	353

OY 302 AAGATTAATTTGCTGTATAGAGTGAATCATGACAACTGATTTTGTGCTGG 361
 Db 354 atgcct--gccacatctccagcttlltaagaaggacaaaagtg999tgaatttcg 411
 OY 362 AGCCGATGCTAAATAATGGAAGCCAGACCTCAAGATTAATGTAAGTGAATCTCC 417
 Db 412 agccaataagaagaagcttgaagccacatlaatgaatcaatcaatgatttc 467

RESULT 2

US-09-385-982-526
 : Sequence 526, Application US/09385982
 : Patent No. 6262334

: GENERAL INFORMATION:

: APPLICANT: ENDEGE, WILSON O., ET AL.
 : TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
 : TITLE OF INVENTION: PRODUCTS: II
 : FILE REFERENCE: CCDNA-260XX
 : CURRENT APPLICATION NUMBER: US/09/385,982
 : CURRENT FILING DATE: 1999-08-30
 : EARLIER APPLICATION NUMBER: 09/328,111
 : EARLIER FILING DATE: 1999-06-08
 : EARLIER APPLICATION NUMBER: 60/117,393
 : EARLIER FILING DATE: 1999-01-27
 : EARLIER APPLICATION NUMBER: 60/098,639
 : EARLIER FILING DATE: 1998-08-31
 : NUMBER OF SEQ ID NOS: 544
 : SOFTWARE: FASTSEQ for Windows Version 3.0
 : SEQ ID NO 526
 : LENGTH: 624
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc.feature
 : LOCATION: (1)...(624)
 : OTHER INFORMATION: n = A,T,C or G
 : US-09-385-982-526

Query Match 12.5%; Score 60; DB 4; Length 624;
 Best Local Similarity 52.0%; Pred. No. 3.4e-06;
 Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

OY 62 TCAGCACTTCCCTGGTGAATCATGTAACGATTATTAAGACAGATGAATTTAAAC 121
 Db 65 tgcctgagctccagcagcaagatgltgaagcagatcgagcgaagcttllcaaga 124
 OY 122 ATTTTGCAGCTGCCGACAACTGCGAGTGTCAATTTCTTGAACGGGTGG 181
 Db 125 agccttgaagcgcagatgtaaaactgtgtagtgaacttcaagcaagtggtg 184
 OY 182 TCCCTGCAAAAGAGTGTTCCTGTTTCCATGAGCTGGCTGAACTTGTCAATCAAAAC 241
 Db 185 gacttgaagaaatgataagccttcttccatccctcctgtaaaagtattccaagtgat 244
 OY 242 AATAACCCACATTTGAGATGTTCAAGAAAGCCAGAGTAACCTATTCTCAAGATCAA 301
 Db 245 attccttgaagatgagtgtg---gatgactgtcagatgtlgttcaagsgtltgaagtc 301
 OY 302 AAGATTAATTTGCTGTATAGAGTGAATCATGAGCAACTGATTTTGTGCTGG 361
 Db 302 atgcctgccaacatccgcttllttaag--aaggacaaaagtg999tgaatttc 358
 OY 362 AGCCGATGCTAAATAATGGAAGCCAGACCTCAAGATTAATGTAAGTGAATCTCC 417
 Db 359 agccaataagaagaagcttgaagccacatlaatgaatcaatcaatgatttc 414

RESULT 3
 US-08-180-371-5
 : Sequence 5, Application US/08180371
 : Patent No. 6254861
 : GENERAL INFORMATION:

: APPLICANT: Choudhury, Chandra
 : TITLE OF INVENTION: Hematopoietic Growth Factor Derived
 : TITLE OF INVENTION: from T Lymphocytes and Methods of Use therefor
 : NUMBER OF SEQUENCES: 18
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Marshall, O'Toole, Gerstein, Murray &
 : ADDRESS: Borun
 : STREET: 6300 Sears Tower, 233 South Wacker Drive
 : CITY: Chicago
 : STATE: Illinois
 : COUNTRY: United States of America
 : ZIP: 60606-6402

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/180,371
 : FILING DATE: 12-JAN-1994
 : CLASSIFICATION: 424

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 906866
 : FILING DATE: 01 July 1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Gruber, Lewis S.
 : REGISTRATION NUMBER: 30,060
 : REFERENCE/DOCKET NUMBER: 27620/31668
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (312) 474-6300
 : TELEFAX: (312) 474-0448
 : TELEX: 25-3856

: INFORMATION FOR SEQ ID NO: 5:

: SEQUENCE CHARACTERISTICS:
 : LENGTH: 630 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cdna
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: 1..630
 : US-08-180-371-5

Query Match 12.5%; Score 60; DB 4; Length 630;
 Best Local Similarity 52.0%; Pred. No. 3.5e-06;
 Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

OY 62 TCAGCACTTCCCTGGTGAATCATGTAACGATTATTAAGACAGATGAATTTAAAC 121
 Db 204 TCGTCAGACTCCAGCAGCAAGATGTAAGACAGATCGAGCAAGACTGTTTCAAGA 263
 OY 122 ATTTTGCAGCTGCCGACAACTGCGAGTGTCAATTTCTTGAACGGGTGG 181
 Db 264 AGCCTTGACCTCCAGGTGATTAACCTGTGATGTGACTGTTCAGCCAGCTGTGG 323
 OY 182 TCCCTGCAAAAGAGTGTTCCTGTTTCCATGAGCTGGCTGAACTTGTCAATCAAAAC 241
 Db 324 GCTTGGCAAAATGATCAAGCCTTCTTCAATCCCTCTCTGTAAGATTTCCAGAGTAT 383
 OY 242 AATAACCCACATTTGAGATGTTCAAGAAAGCCAGAGTAACCTATTCTCAAGATCAA 301
 Db 384 ATTCCTGAAATAGATG---GATGACTGTCAGAGATGTCTTCAAGATGTAAGTCAA 440
 OY 302 AAGATTAATTTGCTGTATAGAGTGAATCATGAGCAACTGATTTTGTGCTGG 361
 Db 441 ATGCATGCCAACATTCAGATTTTAAAG--AAGGACAAAAGGTGGGTGATTTTCTG 497
 OY 362 AGCCGATGCTAAATAATGGAAGCCAGACCTCAAGATTAATGTAAGTGAATCTCC 417
 Db 498 AGCCAAATAGAAAAGCTTGAAGCCACACATTAATGAATTAATGATGATGTTTC 553

RESULT 4
PCT-US92-05707-5
Sequence 5, Application PC/TUS9205707
GENERAL INFORMATION:
APPLICANT: Choudhury, Chandra
TITLE OF INVENTION: Hematopoietic Growth Factor Derived from
T Lymphocytes and Methods of Use Therefor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Bicknell
STREET: Two First National Plaza, 20 South Clark
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/05707
FILING DATE: 19920707
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Gruber, Lewis S.
REGISTRATION NUMBER: 30,060
REFERENCE/DOCKET NUMBER: 27620/30933
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..630
PCT-US92-05707-5

Query Match 12.5%; Score 60; DB 5; Length 630;
Best Local Similarity 52.0%; Pred. No. 3.5e-06;
Matches 185; Conservative 0; Mismatches 165; Indels 6; Gaps 2;

QY 62 TCAGCACTTCCCTGGTGTATCATGTACAGATTATTAAGACGAGTGAATTTAAAC 121
DB 204 TCGCAACACTCCAGCAGCAAGATGTGAAGCAGATGAGAGCAAGCTGCTTTTCAGGA 263
QY 122 ATTTTGAACGTCGCGACAACTCGAGTGTCAATTTCTTCGAAAGGTGG 181
DB 264 AGCCTTGAGCGCTGACAGGTATTAAGTGTAGTGTGACTTCAGCGCAGTGTGG 323
QY 182 TCCCTGCAAAAGATGTTCTGTTTCATGAGCTGGCTGAACCTGTGCATCAAAAC 241
DB 324 GCCTTGCAAAATGATCAAGCTTCTTCATTCCTCTCTCAAAAGATTTCCAGCGAT 383
QY 242 AATACCCACATTTAGATGTTCAAGAAAGCAGAGTAACTTATTTCAAGATCA 301
DB 384 ATTCTTGAGTATGATGTG---GATGACTGTCAGAGTGTCTTCAGAGTGTGAATCA 440
QY 302 AAGATAATTTGCTGTATATGAAGTTCATGAGCAACTGATTTTGTGTTTGG 361
DB 441 ATGATGCGCAACATTTCCAGTTTTAAAG---AAGGACAAAAGGTGGTGAATTTTTCGG 497
QY 362 AGCGATGCTAAAAAATTTGAAAGCAAGACTCAAGAAATTAATTAAGCTGATCTCC 417

DB 498 AGCCAAATAGGAAAAGTTGAGCCAGCCACCATTAATGATTAAGTATCATCATGTTTC 553

RESULT 5
US-08-180-371-17/c
Sequence 17, Application US/08180371
Patent No. 6254861
GENERAL INFORMATION:
APPLICANT: Choudhury, Chandra
TITLE OF INVENTION: Hematopoietic Growth Factor Derived
from T Lymphocytes and Methods of Use Therefor
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/180,371
FILING DATE: 12-JAN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 906866
FILING DATE: 01 July 1992
ATTORNEY/AGENT INFORMATION:
NAME: Gruber, Lewis S.
REGISTRATION NUMBER: 30,060
REFERENCE/DOCKET NUMBER: 27620/31668
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-180-371-17

Query Match 11.8%; Score 56.8; DB 4; Length 630;
Best Local Similarity 51.4%; Pred. No. 2.1e-05;
Matches 183; Conservative 0; Mismatches 167; Indels 6; Gaps 2;

QY 62 TCAGCACTTCCCTGGTGTATCATGTACAGATTATTAAGACGAGTGAATTTAAAC 121
DB 427 TCGTCAAGACTCCAGACGCAAGATGTGAAGCAGATGAGAGCAAGACTGCTTTTCAGGA 368
QY 122 ATTTTGAACGTCGCGACAACTCGAGTGTCAATTTCTTCGAAAGGTGG 181
DB 367 AGCCTTGAGCGCAGCAGGTGATTAAGCTGTAGTGTGACTTCAGCGCAGTGTGG 308
QY 182 TCCCTGCAAAAGATGTTCTGTTTCATGAGCTGGCTGAACCTGTGCATCAAAAC 241
DB 307 GCCTTGCAAAATGATCAAGCTTCTTCATTCCTCTCTCAAAAGATTTCCAGCGAT 248
QY 242 AATACCCACATTTAGATGTTCAAGAAAGCAGAGTAACTTATTTCAAGATCA 301
DB 247 ATTCCTTGAGTATGATGTG---GATGACTGTCAGAGTGTCTTCAGAGTGTGAATCA 191
QY 302 AAGATAATTTGCTGTATATGAAGTTCATGAGCAACTGATTTTGTGTTTGG 361


```

TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1461 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 54..617
PCT-US95-04258-4

Query Match          9.6%; Score 46.4; DB 5; Length 1461;
Best Local Similarity 70.5%; Pred. NO. 0.0081;
Matches 62; Conservative 0; Mismatches 26; Indels 0; Gaps 0

QY 394 AAGATTAAATGTACCTGATCTCCCAAGCCAAATACACTTGTGACATTTGAAAAGGCCAA 453
      || ||||| || || || || || || || || || || || || || || || || || ||
Db 1347 AAAAATTAAAGAGAGCCCAATTAATAATAAAAAATACCTTTCGTCTGTAATAAAAAA 1406

QY 454 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1407 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434

```

RESULT 10
 US-08-232-463-14/c
 : Sequence 14, Application US/08232463
 Patent No. 5670367
 : GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEFFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single

Db 1441 TGACATTAAGCTCTCTTGTGACACATCAATTAAGCTTGTCTGATGTAATCTTGA 1500
QY 375 AATTTGGAACCCAGACTCAAGATTAATTAAGCTGATCTCCAGGCAAAATACACTTG 434
Db 1501 CATGTAAATCACACATTTAAAGATATATCATTCGAGTTCGCTTAAATAAAAA 1560
QY 435 TGACATTTGAAAGCAAAAAA 478
Db 1561 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1604

RESULT 13

US-08-514-014-1
; Sequence 1, Application US/08514014
; Patent No. 5707829
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Kelleher, Kerry
; APPLICANT: Carlin, McKeough
; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/514,014
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G16000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2209 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 38..1447
; US-08-514-014-1

Query Match 9.4%; Score 45.4; DB 1; Length 2209;
Best Local Similarity 67.4%; Pred. No. 0.015;
Matches 64; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 387 AAGACTCAAGATTAATGTAAGCTGATCTCCAGGCAAAATACACTTGTGACATTTGAAA 446
Db 2112 AAGGCTTAATGATTTATGTAACCATATATGTCATAAAACACACTTTTGAAGCAAAAA 2171
QY 447 AGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 481
Db 2172 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2206

RESULT 14

US-08-833-823-1
; Sequence 1, Application US/08833823
; Patent No. 5969093
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Kelleher, Kerry
; APPLICANT: Carlin, McKeough
; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,823
; FILING DATE: 10-APR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/514,014
; FILING DATE: 11-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G16000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2209 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 38..1447
; US-08-833-823-1

Query Match 9.4%; Score 45.4; DB 2; Length 2209;
Best Local Similarity 67.4%; Pred. No. 0.015;
Matches 64; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 387 AAGACTCAAGATTAATGTAAGCTGATCTCCAGGCAAAATACACTTGTGACATTTGAAA 446
Db 2112 AAGGCTTAATGATTTATGTAACCATATATGTCATAAAACACACTTTTGAAGCAAAAA 2171
QY 447 AGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 481
Db 2172 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2206

RESULT 15

US-08-813-150-1
; Sequence 1, Application US/08813150
; Patent No. 6089229
; GENERAL INFORMATION:
; APPLICANT: Mueller, Christopher
; APPLICANT: Lebecque, Serge J.E.

APPLICANT: Liu, Yong-Jun
APPLICANT: Dowling, Lynette M.
APPLICANT: Huffine, Constance M.
APPLICANT: Gorman, Daniel M.
TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;
RELATED REAGENTS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/813,150
FILING DATE: 07-MAR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: SF0693
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-852-9196
TELEFAX: 650-496-1200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 61..1470
US-08-813-150-1

Query Match 9.4%; Score 45.4; DB 3; Length 2280;
Best Local Similarity 67.4%; Pred. No. 0.016;
Matches 64; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 387 AAGACTCAAGAAATTAATGTAAGCTGATCTCCAGGCAAAATATCACTTGACATTGAA 446
DB 2134 AAGGCTTAATGATTATGAACCAATATGTCATATAAACACACACTTTTGAGGCAAAA 2193
QY 447 AGGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 481
DB 2194 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2228

Search completed: June 8, 2002, 03:21:07
Job time: 3947 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2002, 02:09:55 ; Search time 1800.24 Seconds
(without alignments)
3606.206 Million cell updates/sec

Title: US-09-746-783-18

Perfect score: 481
Sequence: 1 GGATACGTGATATAATAGCA.....AAAAAAAAAAAAAAAAAAAA 481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estf2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149.6	31.1	399	9	A1188241 qd67305.x
2	144.6	30.1	619	11	AK015240 Mus muscu
3	144.2	30.0	491	11	AK006683 Mus muscu
4	141.4	29.4	526	9	AV209695 AV209695
5	126.4	26.3	344	9	A1764117 UI-R-YO-a
6	118	24.5	245	9	AV045293 AV045293
7	109.4	22.7	266	9	AV269460 AV269460
8	96	20.0	666	12	AG084227 Pan trogl
9	85	17.7	286	9	AV269811 AV269811
10	83.8	17.4	384	10	BF149520 BF149520
11	83.6	17.4	428	9	A1885805 A1885805
12	75.2	15.6	502	12	AQ351522 RC111-11
13	73.4	15.3	907	10	BF028142 BF028142
14	72.8	15.1	273	9	AV267588 AV267588
15	68.6	14.3	846	10	BF028467 BF028467
16	67.6	14.1	460	10	Z71861 CHEST231 Go
17	67.4	14.0	849	10	BF210792 BF210792

18	67.2	14.0	424	10	W87705 W87705
19	66.4	13.8	468	10	W95350 W95350
20	65.4	13.6	459	10	BF109086 BF109086
21	65.2	13.6	555	9	A1815810 A1815810
22	65.2	13.6	569	10	BG567720 BG567720
23	64.8	13.5	763	10	BE787457 BE787457
24	64.6	13.4	546	10	BF970488 BF970488
25	64	13.3	455	10	W69590 W69590
26	63.8	13.3	426	10	BG937432 BG937432
27	63.8	13.3	427	10	BM429932 BM429932
28	63.8	13.3	424	10	BM433322 BM433322
29	63.8	13.3	516	9	AV663283 AV663283
30	63.8	13.3	561	9	AV593496 AV593496
31	63.8	13.3	603	10	BG027655 BG027655
32	63.4	13.2	560	10	BF246611 BF246611
33	63.2	13.1	393	9	AJ275295 AJ275295
34	63.2	13.1	455	9	AW429592 AW429592
35	63.2	13.1	472	10	B1359967 B1359967
36	63.2	13.1	522	9	A1620253 A1620253
37	63.2	13.1	544	10	BF218303 BF218303
38	63.2	13.1	575	10	B1346139 B1346139
39	63	13.1	501	9	AA810660 AA810660
40	63	13.1	546	10	BE967335 BE967335
41	62.8	13.1	397	10	BE440038 BE440038
42	62.8	13.1	458	9	AV593495 AV593495
43	62.6	13.0	485	10	N77546 N77546
44	62.4	13.0	461	10	W76348 W76348
45	62.4	13.0	503	9	AA045052 AA045052

ALIGNMENTS

RESULT 1
A1188241/c
LOCUS
DEFINITION
qd67305.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1734537
3' similar to SW:THIO_HUMAN P10599 THIOREDOXIN ;, mRNA sequence.

ACCESSION
A1188241
VERSION
A1188241.1 GI:3739450
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: gcgaps-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.C.E. Consortium/LNLN at:
www.bio.linn.gov/bdrp/image/image.html
Insert length: 523 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 347.
Location/Qualifiers
1. 399

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1734537"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT773D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech laboratories


```

121 GATTAAAGACATGACGTGAACATGTAACGATGCGAAGCAAGCTTGCG 180
154 TGGTCAATTTCTTCGAAGCGTGCTCCGCAAAAGATGTTCTGTTTCATG 213
181 TGGTAGAGTTTTCAGCAAGTGGTGGCCCTGCAAAACAAATGCTCTGTTTCCAGG 240
214 AGCTG----- 218
241 CAATGCTTTAAATACCAAAATGTCAAGTTTGTCTCAGGTGATGTGACTCATTAAG 300
219 -----GCTGAACCTTGTCACATCAAAACATACCCACATTCAGATGTTCAAGAAAGCC 273
301 AATTAGCTGACATGTGACATGCAATGCTACCCACATTCACATGCTTCAAGTACACCC 360
274 AGAAGTACCCCTGCTTCGCAAGATCAAAAGATTAATTTGCTTTATAGAGTGAATTC 333
361 AAAAGTCTCCATCTTCCTCAGACTCAAAAGAGTGTCTGCTCCTCAGAGTGGACCGA 420
334 TGAACAACCTGATTTTGTGAGTTTGTGAGCCGATGCTAAATAATTTGGAAGCAAGCTC 393
421 AGAGCAAGATGATTTTGTGAGTTTGTGAGCCGATGCTAAATAATTTGGAAGCAAGCTC 480
394 AAGATTAATGTAAGCTGATCTCGAA 419
481 AAGAGTAATGTAAGCTGATCTCGAA 506

```

```

RESULT 5
AT764117/c 344 bp mRNA linear EST 25-JUN-1999
LOCUS
DEFINITION
UT-R-Y0-acr-a-07-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
UT-R-Y0-acr-a-07-0-UI 3', mRNA sequence.

```

```

ACCESSION AT764117 GI:5210052
VERSION AT764117
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Rodentia; Scuriognathi; Muridae; Murinae;
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE 1 (bases 1 to 344)
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLES Normalization and subtraction: two approaches to facilitate gene
discovery.
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

```

```

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NCI site
and the oligo-dT track served to verify it as a clone from the
normalized Eye library cDNA library preparation: M.B. Soares Lab
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLY-A-yes.

```

```

FEATURES
Source Location/Qualifiers
1..344

```

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-acr-a-07-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

```

```

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)
TAG_LIB=UI-R-Y0
TAG_TISSUE=eye
TAG_SEQ=CATTTG"
BASE COUNT 90 a 70 c 69 g 115 t
ORIGIN

```

```

Query Match 26.3%; Score 126.4; DB 9; Length 344;
Best Local Similarity 69.4%; Pred. No. 3.1e-07;
Matches 172; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```

```

210 CATGAGCTGGCTGAACTTGCATCAAAACATACCCACATTCATGTTCAAGAA 269
248 CAGGATGAGCTGAACTTGCATCAAAACATACCCACATTCATGTTCAAGAA 189
270 AGCCGAGGATGAGCTTGTGAGTTTGTGAGCCGATGCTAAATAATTTGGAAGTGA 329
188 TCCGCAAGAGTCACTCCATTCAGACTCAAAACATACCTGTTGTTCAAGAGTGA 129
330 TTCATGAGCAACCTGATTTTGTGAGTTTGTGAGCCGATGCTAAATAATTTGGAAGTGA 389
128 CCAGGAGCAAGAGATTTTGAATTTTGAAGCCGATGTTGAATAATTTGGAAGAG 69
390 ACTCAAGATTAATGTAAGCTGATCTCAAGGCAAAATACCTTGTGATTTGAAAGG 449
68 ATTCAAGAGCTAATGTAAGCTGATCTCAAGGCAAAATACCTTGTGATTTGAAAGG 9
450 CAAAAAAA 457
8 AAAAAAAA 1

```

```

RESULT 6
AV045293 245 bp mRNA linear EST 23-NOV-1999
LOCUS
DEFINITION AV045293 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
1700042D21, mRNA sequence.
ACCESSION AV045293
VERSION AV045293.2 GI:4864958
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.

```

```

REFERENCE 1 (bases 1 to 245)
AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara,
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Iwasa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shitaki, T., Sogabe, Y., Suganara,
Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,

```


RESULT	9
AV269811	
LOCUS	
DEFINITION	AV269811 RIKEN full-length enriched, adult male testis (DH10B) Mus
	296 bp mRNA linear EST 05-NOV-1995
	AV269811

musculus cdna clone 4930545G19 3', mRNA sequence.
AV269811
AV269811.1 GI:6257848
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 296)
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,
C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakanuma, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata
Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
Suzuki, H., Takahashi, F., Tateo, M., Tomiyaga, N., Tsunoda, Y.,
Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,
Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al. 1999)
Unpublished (1999)
Contact: Yoshinobu Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki, M., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki
Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.riken.go.jp) for
further details.

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930545G19"
/clone_1lb="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGACAGGAGGAGGATCCACAGACCTTTTTTTTTTTTTTAA 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGACAGGAGGATTCGCGATTAAATTATATCCCCCCCCCC 3']. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified plasmid pUC(+), after bulk excision from lambda
phage. Cloning sites: 5' end: SalI; 3' end: BamHI."

```

ORIGIN	BASE COUNT	90 a	67 c	63 g	76 t
Query Match	17.7%	Score 85;	DB 9;	Length 296;	
Best Local Similarity	67.5%;	Pred. No. 0.049;			
Matches 137;	Conservative 0;	Mismatches 60;	Indels 6;	Gaps 1;	
OY	257	GATGTTCAAGAAACCCAGAGTAACTACCTATTTCACAGATCAAAAGCATATTGCTG	316		
Db	1	GATCTCAACTACACCCAAAAGTTACCTCAATTTTCAGACACGCCAAGAGGAGTGTGATG	60		
OY	317	TTATAGAGTGATTCATGACGACACCTGATTTTGATGTTTGAGAGCGGATGCTAAAAA	376		
Db	61	CCTCAGAGTGCACCCGAGAGGACGACATGATTTTAAAGTCAATGAGGCTGATGCTAAACA	120		
OY	377	ATTGAGCCAGCAACTCAAGAAATATGATTAAGCTGATCTCCAGGCCAAATACACTGTG	436		
Db	121	ACTGAAAAAGAGATTCAAGACTAATGTAAAGGCCACCC-----AATATACTGTG	174		
OY	437	ACATTGAAAGGCCAAAAAA 459			
Db	175	ACATTTCCTAAAGCAGAGCTAA 197			
RESULT 10					
LOCUS	BF149520				
DEFINITION	BF149520	384 bp	mRNA	linear	EST 29-DEC-2000
KEYWORDS	BF149520	uy66a02.y1	Mccarrey Eddy	round spermatid	Mus musculus cDNA clone
ACCESSION	BF149520	IMAGE:3664490	5'	similar to	SW:THIO_RABIT P08628 THIOREDOXIN. ;
VERSION	BF149520	GI:11030915			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 384)				
TITLE	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.				
JOURNAL	The WashU-NCI Mouse EST Project 1999				
COMMENT	Unpublished (1999)				
	Washington University School of Medicine				
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA				
	Tel: 314 286 1800				
	Fax: 314 286 1810				
	Email: mouseest@wustl.edu				
	This clone is available royalty-free through LNL; contact the				
	IMAGE Consortium (info@image.lnl.gov) for further information.				
	MC1:1425258				
	Seq primer: Primer name ambiguous				
	High quality sequence stop: 368.				
FEATURES	location/Qualifiers				
source	1. 384				
	/organism="Mus musculus"				
	/strain="CD-1"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:3664490"				
	/clone_id="Mccarrey Eddy round spermatid"				
	/sex="male"				
	/tissue_type="round spermatids, pooled from multiple mice"				
	/dev_stage="60 day"				
	/lab_host="DHI08 (phage-resistant)"				
	/note="Organ: testis; Vector: pBluescript SK+ (Stratagene				
	[5'-(G)10-ACTGACGTCGAGTGTGTGTGT-3'] and directionally				
	cloned using 5' linkers 5'-ATTGGCAGCAG-3' and				
	5'-CTGTCGCCG-3'. Size selection of >400bp material gives				

BASE COUNT	117 a	86 c	90 g	91 t
ORIGIN	average insert size ranging from 1-2 kb. Library was mass excised (from Lambda-UniZAP-Xr) and resulting single-stranded phagemids were prepried and transformed into DH10B. Library contains 98.5% recombinants. References: J. Androl. 20:635-639 and Gene 25:263-269. Library constructed and bloned by J. McCarrey, Ph.D. (Southwest Foundation for Biomedical Research, Dept. of Genetics); excision done by E.M. Eddy, Ph.D. (National Institutes of Health, National Institute of Environmental Health Sciences). Original lambda-based library is available through ATCC, catalog #63423."			
Query Match	17.4%	Score 83.8;	DB 10;	Length 384;
Best Local Similarity	76.3%;	Pred. No. 0.059;		
Matches 103;	Conservative	0;	Mismatches 32;	Indels 0;
Gaps 0;				
OY 213	GAGTGGCTGAACACTGTCACATCAAAACATACCACATTCAGATGTCACAGAAAAGC	272		
Db 250	GAAATGACGTGAACATGTGACATGCATCATGTCACCATTCACAGATGTCACAGTACACC	309		
OY 273	CAGAAGTACACCTTTCTCAGAGATCAAAAGATTAATTGCTGTATAGAAGTGCATTC	332		
Db 310	CAAAAGTCACATCCATTTCTCAAGACTCAAAAGAGTACTGTCCTGCCTCAAGATGCACCG	369		
OY 333	ATGAGCAACCTGATT 347			
Db 370	AAGACCAAGATGATT 384			
RESULT 11				
AI885805/c				
LOCUS	AI885805	428 bp	mRNA	linear
DEFINITION	w161g04.x1 NCI_CGAP_Brn25 Homo sapiens CDNA clone IMAGE:2429430 3'			
ACCESSION	AI885805			
VERSION	AI885805.1	GI:5590969		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 428)			
JOURNAL	NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
COMMENT	National Cancer Institute / National Institute of Neurological			
	Disorders and Stroke, Brain Tumor Genome Anatomy Project			
	(CGAP/BRGAP), Tumor Gene Index			
	Unpublished (1998)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgaps-remail.nih.gov			
	Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,			
	Ph.D.			
	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima			
	Bonaldo, Ph.D.			
	CDNA Library Arrayed by: Greg Lennon, Ph.D.			
	DNA Sequencing by: Washington University Genome Sequencing Center			
	Clone distribution: NCI-CGAP clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LINL at:			
	www-bio.lnll.gov/db/ftp/image/image.html			
	Seq primer: -400p from Gibco			
	High quality sequence stop: 404.			
FEATURES	location/Qualifiers			
SOURCE	1. 428			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:2429430"			
	/clone_1bp="NCI_CGAP_Brn25"			
	/tissue_type="anaplastic oligodendroglioma"			
	/lab_host="DH10B"			
	/note="Organ: brain; Vector: pF773D-Pac (Pharmacia) with a			
	modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st			

The Not I and Eco RI sites of the modified pT713 vector. Library is normalized, and was constructed by Benito Soares and M.Fátima Bonafide.

BASE COUNT	123 a	87 c	79 g	139 t
ORIGIN				

Query Match	17.4%;	Score	83.6;	DB	9;	Length	428;
Best Local Similarity	59.1%;	Pred. No.	0.059;				
Matches	143;	Conservative	0;	Mismatches	99;	Indels	0;
				Gaps	0;		

RESULT 12	
A0351522/c	
LOCUS	A0351522
DEFINITION	502 bp DNA linear GSS 07-MAY-1998
	RPcII-111D7.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-111D7,
	DNA sequence.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

```

/zb_xref="GDB:7542318"
/zb_xref="taxon:9606"
/clone="RPC1-11-111D7"
/clone_11b="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/notes="Vector: pBACe3.6; Site_1: ECOR1, Site_2: ECOR1,
RPC111 Human Male BAC Library"

```

Query Match	15.68;	Score 75.2;	DB 12;	Length 502;
Best Local Similarity	63.08;	Pred. No. 0.59;		
Matches 116; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0;

RESULT	13
BF028142	
LOCUS	BF028142
DEFINITION	601764903031 NIH_MGC_53 Homo sapiens CDNA clone IMAGE:3997035 , mRNA sequence.

```
FEATURES      Location/Qualifiers
source        1..502
              /organism="Homo sapiens"
```


VERSION BF028467.1 GI:10736179
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 846)
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM847 row: k column: 17
High quality sequence stop: 505.

FEATURES

1, 846
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:396040"
/clone_lib="NIH-MGC_53"
/tissue_type="Cardioma, cell line"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: bladder; Vector: pDNR-LIB (Clontech); Site_1: Sfil (ggccgctcgcc); Site_2: Sfil (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCGCGGAGCGCCGACATG-3' and 3' adaptor (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
BASE COUNT 256 a 199 c 210 g 180 t 1 others
ORIGIN

Query Match 14.3%; Score 68.6; DB 10; Length 846;
Best Local Similarity 51.1%; Pred. No. 2.8; Mismatches 199; Indels 6; Gaps 2;
Matches 214; Conservative 0;

```

QY 62 TCAGCACTTCCTGCTGTAATCATGTATCAATTAAGACGAAATGAATTAAMC 121
   || || || || || || || || || || || || || || || || || || ||
DB 30 TCGTCAGACTCCAGCAAGCAAGATGCTGAAGACAGATCGAGCAAGACTCTTTCAG 89
   || || || || || || || || || || || || || || || || || || ||
QY 122 ATTTTGAACGCTGCCGACCAACTCGCAGTGGTTCAATTTCTTCGAAACGGTGG 181
   || || || || || || || || || || || || || || || || || || ||
DB 90 ACCCTTGGACGGCTGATTAACCTTGATGATGACTTCTCAAGCCAGTGGTGG 149
   || || || || || || || || || || || || || || || || || || ||
QY 182 TCCCTGCAAAAGATGTTCCGTTTCATGAGCTGGCGAAACTTGTCAACATCAAA 241
   || || || || || || || || || || || || || || || || || || ||
DB 150 GCCTTCAAAATGATCAAGCTTCTTCTCTCTCTGAAAAGATTCACACGTGAT 209
   || || || || || || || || || || || || || || || || || || ||
QY 242 AATACCAATTTCAAGATGTTCAAGAAAGCAAGATTAACCTATTTCAAGAAATCA 301
   || || || || || || || || || || || || || || || || || || ||
DB 210 ATTCTTGAAGTAGATGTG---GATGACGTGTAGATGTTGCTTCAGAGTGTGAA 266
   || || || || || || || || || || || || || || || || || || ||
QY 302 AAGAAATATTTGCTGTTATAGAAAGTATTCATGAGCAACCTGATTTTGAAGTT 361
   || || || || || || || || || || || || || || || || || || ||
DB 267 ATGCATGCCAACATTCAGTTTTTTTAAG--AAGGGCAAAAGGTGGTGAATTTCT 323
   || || || || || || || || || || || || || || || || || || ||
QY 362 AGCCGATGCTAATAAATTTGAAGCAAGACTCAAGATTAATGAAGCTGATCCCAAG 421
   || || || || || || || || || || || || || || || || || || ||
DB 324 AGCCATTAAGAAACCTTGAAGCCACCATTAATGAATTAATCAATGTTTCTGAA 383
   || || || || || || || || || || || || || || || || || || ||
QY 422 CAAATATCACTTGTGACATTTGAAAGGCAAAAAAAAAAAAAAAAAAAAAAAAAAA 480
   || || || || || || || || || || || || || || || || || || ||

```

DB 384 ACATAACCAAGCCATGGCTATTAAACTGTAAATTTTTTTTAACTTACAAAAATATAAA 442

Search completed: June 8, 2002, 02:48:05
Job time: 2290 sec

Sat Jun 8 10:14:12 2002

us-09-746-783-18.rst

Page 12

> O <
01 10 Intelligenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file Seq19.res made by tport on Mon 24 Jun 102 8:34:12-PDT.

Query sequence being compared: US-09-746-783-19 (1-107)
Number of sequences searched: 1
Number of scores above cutoff: 1

Results of the initial comparison of US-09-746-783-19 (1-107) with:
File : p10599.pep

```

100-  N -
      U -
      M -
      B -
      E -
      R -
      O -
      F 10-
      S -
      E 5-
      Q -
      D -
      E -
      N -
      C -
      E -
      S 0-
SCORE 0 4 8 13 17 21 25 30 34 38
STDEV

```

PARAMETERS

Similarity matrix PAM-150 K-tuple 1
Threshold level of sim. 16%
Mismatch penalty 1 Joining penalty 20
Gap penalty 5.00 Window size 107
Gap size penalty 0.05
Cutoff score 1
Randomization group 0

SEARCH STATISTICS

Scores: Mean 38 Median 0 Standard Deviation 0.00
Times: CPU 00:00:00.00 Total Elapsed 00:00:00.00
Number of residues: 105
Number of sequences searched: 1
Number of scores above cutoff: 1

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Opt. Frame
---------------	-------------	--------	-------	------------

1. thiohuman	TOIG of: thio_human	check: 3	105	38 61 0.00 0
--------------	---------------------	----------	-----	--------------

1. US-09-746-783-19 (1-107)
thiohuman TOIG of: thio_human check: 3392 from: 1 to: 105

Initial Score = 38 Optimized Score = 61 Significance = 0.00
Residue Identity = 31% Matches = 34 Mismatches = 65
Gaps = 2 Conservative Substitutions = 6

```

X      10      20      30      40      50      60      70
MVOIIRKDTNEFKTFLTAAGHKLAVVQSSKRCGPCKRMFPVPHLEAETCHIKTIPTFQMFKKSOKVTLFSRI
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
MVKQIESKTAFQELDAAGDKLVVWDFSATWCGPCMKIKPFPHSLSEKY--SNVIFLEVVDVDDCODVASCE
X      10      20      30      40      50      60      70
KRIICYSRSGFMSNLIFEECGADAKKLEAKTQELM
:      :      :      :      :      :      :      :      :      :      :      :
VKCMPTFOFPFKKGQKVGEGESGANKKELEATINELY
80      90      100      X
80      90      100      X

```


CAGCTACAGTATCCACGTAACCAACCAACGAGGATTTTCACTACACAGCTCCCTGGTATCATGATACAGANT	30	40	50	60	70	80	90	
CACCGGATTACTCCCGGTCACCGTTACTACGCTCTT--TGTGGGTTCACGCGCTGGGGGGCCGGCGCCG	100	110	120	130	140	150	160	
210	220	230	240	250	260	270		
TATTAAAGACAGCATGTAATTTAAACAATTTTATGACAGCTGCCGAGACAAACTCCAGTGGTTCATATTTTC	100	110	120	130	140	150	160	
TGGCGGCCCCCTTTTCGAGGAATCCAGCCCGCTGGGGGTCCCAATCTGAGCGTGGGGGTGTTGATTACAG	280	290	300	310	320	330	340	
170	180	190	200	210	220	230		
TTGCAACCGTGTGTGTCCTCC--TGCAAAAGAGATGTTCCTGTTTCCATGATGAGCTGGCTG--AAACTTCTAC								
350	360	370	380	390	400	410		
GCCCGCGGAGCATTCCCGAGTGACCCCGGAGGGAGCTTGTCCTCTGGGGGTGACTGCTCGTCCGCCG								
240	250	260	270	280	290	300		
ATCAAAACAATACCCACATTTTCAGTGTTCAGAAGAAAGCAAGGATACCTATTCTCAGAAATCAAGAAG								
310	320	330	340	350	360			
ATTAATTTGCTGTTAT-----AGAAAGTGATTTATGACCAACTGAT-----TTTGAAGTTTGTGGAGCCG								
CGAAGCCCTGTGAGGCTCCGCCACGCTTTGCCGCGCCGCCCGCCGATATCCCGGCTGCAGAGCTCCAGGGCGG	430	440	450	460	470	480		
500	510	520	530	540	550	560		
370	380	390	400	410	420	430		
ATGCTTAAAAAATTTGAGAGCAAGCACTCAAGAATATGTATGTGATTCACAAAGCAAAATATGACTTTGTGAC								
GGGTGCGCGGGGGCCGACGAGCGGGCGGGAGGTTCGGCCGGCGCGCGCCCGCCGACACGACCGCCGG	570	580	590	600	610	620	630	
440	450	460	470	480				
ATTGCAAAAGCGCAAAAAAAAAAAAAAAAAAAAAAAAAA								
GGCTGCCAGTTTATTAAGGAGAGAGCAAGCAGCATCTTTGAAGCTGTGTTGGGTCTTTGATACCATTTTC	640	650	660	670	680	690	700	
CATGGTCTCTTACAGCCGCTC	710							

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 04:29:56 ; Search time 36.79 Seconds

(without alignments)
112.612 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562
Sequence: 1 MVOIKDFNEFKFLTAAGH.....IFEFGADAKLEAKTQELM 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	185	32.9	104	THIO_HUMAN	P10599 Homo sapien
2	185	32.9	104	THIO_SHEEP	P50413 ovis aries
3	184	32.7	104	THIO_RABIT	P08628 corycolagus
4	182	32.4	104	THIO_BOVIN	O97680 bos taurus
5	180	32.0	104	THIO_MOUSE	P10638 mus musculu
6	179	31.9	104	THIO_RAT	P11232 ratulus norv
7	177	31.5	104	THIO_MACMU	P29451 macaca mola
8	160	28.5	104	THIO_CHICK	P08629 gallus gall
9	124	22.1	106	THIO_COPCM	O94002 coprinus co
10	118	21.0	289	TXNL_HUMAN	O43396 homo sapien
11	111	19.8	102	THIO_CHPN	O92795 chlamydia p
12	110	19.6	102	TRX2_SCHPO	O14463 schizosacch
13	108	19.2	102	TRX1_SCHPO	P58465 schizosacch
14	106.5	19.0	182	THIF_PEA	P29450 pisum sativ
15	103.5	18.4	190	THIF_SPIOI	P09856 spinacia ol
16	101.5	18.1	107	THIO_DROME	P47938 drosophila
17	101.5	18.1	185	THF2_ARATH	O6Xf8n arabidopsis
18	100.5	17.9	133	THH2_ARATH	O38879 arabidopsis
19	100.5	17.9	182	THIF_BRANA	O48897 brassica na
20	99	17.6	105	THI1_DICDI	P29445 dictyosteli
21	98.5	17.5	178	THF1_ARATH	O9Xf8n arabidopsis
22	98	17.4	102	THIO_CVAME	O32022 cyanidiosi
23	98	17.4	104	THI3_DICDI	P29447 dictyosteli
24	97.5	17.3	103	TRX1_YEAST	P22803 saccharomyc
25	96.5	17.2	191	THIF_MESCR	O61332 mesembryant
26	95	16.9	102	THIO_CHLMU	O9Pjk3 chlamydia m
27	94	16.7	112	THIO_CHLRE	O84544 chlamydia t
28	92.5	16.5	112	THIO_CHLRE	P80028 chlamydomon
29	92	16.4	102	THIO_CHLPS	P52227 chlamydia p
30	92	16.4	127	THIO_NEUCR	P42115 neurospora
31	90.5	16.1	122	THI1_ORYSA	O42443 oryza sativ
32	89	15.8	88	THI2_DICDI	P29446 dictyosteli
33	89	15.8	126	THI1_WHEAT	O64394 triticum ae

34	88.5	15.7	107	THIO_CVACA	P37395 cyanidium c
35	86	15.3	118	THH5_ARATH	O39241 arabidopsis
36	86	15.3	175	THIM_WHEAT	O92p21 triticum ae
37	85.5	15.2	106	THIO_PENCH	P34723 penicillium
38	85.5	15.2	114	THH1_ARATH	P29448 arabidopsis
39	85.5	15.2	172	THIM_PEA	P48384 pisum sativ
40	84.5	15.0	156	DYL5_CHLRE	O39592 chlamydomon
41	84.5	15.0	519	PDAS_HUMAN	O14554 homo sapien
42	84	14.9	102	TRX2_YEAST	P22217 saccharomyc
43	84	14.9	109	THIO_EMENT	P29429 emericocella
44	83.5	14.9	115	THIO_MCTU	P52228 mycobacteri
45	83	14.8	115	THIO_CABEL	O09433 caenorhabdi

ALIGNMENTS

RESULT	1	THIO_HUMAN	STANDARD:	PRT:	104 AA.
ID	THIO_HUMAN	P10599:			
AC	01-JUL-1989 (Rel. 11, Created)				
DT	01-DEC-1992 (Rel. 24, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Thioredoxin (ATL-derived factor) (ADF) (Surface associated sulphhydryl protein) (SASP).				
GN	TXN OR TRX OR TRX.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=91340156; PubMed=1874447;				
RT	Tonissen K.F., Wells J.R.E.;				
RT	"Isolation and characterization of human thioredoxin-encoding genes.";				
RL	Gene 102:221-228(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=89008454; PubMed=3170595;				
RA	Wollman E.E., D'Aurion L., Rinsky L., Shaw A., Jacquot J.-P.,				
RT	Wingfield P., Graber P., Dessars F.;				
RT	"Cloning and expression of a cDNA for human thioredoxin.";				
RL	J. Biol. Chem. 263:15506-15512(1988).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=89251607; PubMed=2785919;				
RA	Tagaya Y., Maeda K.-I., Mitsui A., Kudo N., Matsui H., Hamuro J.,				
RT	Brown N., Arai K.-I., Yokota T., Wakasugi H., Yodoi J.;				
RT	"ATL-derived factor (ADF), an IL-2 receptor/Trac inducer homologous to				
RT	thioredoxin; possible involvement of dithiol-reduction in the IL-2				
RT	receptor induction.";				
RL	EMBO J. 8:757-764(1989).				
RN	[4]				
RP	SEQUENCE OF 1-14.				
RA	MEDLINE=91151337; PubMed=1998498;				
RT	Martin H., Dean M.;				
RT	"Identification of a thioredoxin-related protein associated with				
RL	plasma membranes.";				
RN	Biochem. Biophys. Res. Commun. 175:123-128(1991).				
RN	[5]				
RP	STRUCTURE BY NMR.				
RA	MEDLINE=90057393; PubMed=2684271;				
RA	Forman-Kay J.D., Clore G.M., Driscoll P.C., Wingfield P.,				
RT	Richards F.M., Gronenborn A.M.;				
RT	"A proton nuclear magnetic resonance assignment and secondary				
RL	structure determination of recombinant human thioredoxin.";				
RN	Biochemistry 28:7088-7097(1989).				
RN	[6]				
RP	STRUCTURE BY NMR.				
RA	MEDLINE=91159399; PubMed=2001356;				
RA	Forman-Kay J.D., Clore G.M., Wingfield P., Gronenborn A.M.;				
RT	"High-resolution three-dimensional structure of reduced recombinant				

```

RT      human thiorodoxin in solution." ;
RL      Biochemistry 30:2685-2698(1991).
RN
RP      STRUCTURE BY NMR.
RX      MEDLINE=95006318; PubMed=7922028;
RA      Qin J., Clore G.M., Gronenborn A.M.;
RT      "The high-resolution three-dimensional solution structures of the
RL      oxidized and reduced states of human thiorodoxin." ;
RN      Structure 2:503-522(1994).
RP      STRUCTURE BY NMR.
RX      MEDLINE=96347359; PubMed=8736558;
RA      Qin J., Clore G.M., Kennedy W.P., Kuszewski J., Gronenborn A.M.;
RT      "The solution structure of human thiorodoxin complexed with its
RL      target from Ref-1 reveals peptide chain reversal." ;
RN      Structure 4:613-620(1996).
RP      X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RX      MEDLINE=96399719; PubMed=8805557;
RA      Welchesel A., Gasdaska J.R., Powis G., Montfort W.R.;
RT      "Crystal structures of reduced, oxidized, and mutated human
RL      thiorodoxins: evidence for a regulatory homodimer." ;
RN      Structure 4:735-751(1996).
RP      X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF MUTANT ASN-60.
RX      MEDLINE=968039128; PubMed=9369469;
RA      Andersen J.F., Sanders D.A., Gasdaska J.R., Welchesel A., Powis G.,
RT      Montfort W.R.;
RL      "Human thiorodoxin homodimers: regulation by pH, role of aspartate
RT      60, and crystal structure of the aspartate 60 --> asparagine
RL      mutant." ;
RN      Biochemistry 36:13979-13988(1997).
RP      ACTIVITY.
RX      MEDLINE=91097576; PubMed=2176490;
RA      Jacquot J.-P., de Lamotte F., Fontecay M., Schuermann P.,
RT      Decoligantes P., Miglinac-Maslow M., Wollman E.;
RL      "Human thiorodoxin reactivity-structure/function relationship." ;
RN      Biochem. Biophys. Res. Commun. 173:1375-1381(1990).
RP      -I- FUNCTION: THIORODOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
RT      THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC      TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC      -I- FUNCTION: AID AUGMENTS THE EXPRESSION OF THE INTERLEUKIN-2
CC      RECEPTOR TAC (IL2R/P55).
CC      -I- SIMILARITY: BELONGS TO THE THIORODOXIN FAMILY
CC
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; X77584; CAA54687.1; -
DR      EMBL; X54539; CAA38410.1; -
DR      EMBL; X54540; CAA38410.1; JOINED.
DR      EMBL; X54541; CAA38410.1; JOINED.
DR      EMBL; J04026; AAA74596.1; -
DR      PIR; A31993; A31993.
DR      PIR; S04106; S04106.
DR      PIR; JH0568; JH0568.
DR      PDB; 3TRX; 15-JAN-93.
DR      PDB; 4TRX; 15-JAN-93.
DR      PDB; 1TRS; 30-SEP-94.
DR      PDB; 1TRU; 30-SEP-94.
DR      PDB; 1TRV; 30-SEP-94.
DR      PDB; 1TRW; 30-SEP-94.
DR      PDB; 1AIU; 07-JUL-97.
DR      PDB; 1ERT; 14-OCT-96.
DR      PDB; 1ERU; 01-AUG-96.
DR      PDB; 1ERW; 14-OCT-96.
DR      PDB; 1ERW; 14-OCT-96.

```

DR	PDB:	1CQG;	01-AUG-96.
DR	PDB:	1COH;	01-AUG-96.
DR	PDB:	1MDI;	03-JUN-95.
DR	PDB:	1MDJ;	20-JUL-95.
DR	PDB:	1MDK;	20-JUL-95.
DR	PDB:	1AUC;	25-FEB-98.
DR	SWISS-2DPAGE:	P10599;	HUMAN.
DR	Aarhus/Ghent-2DPAGE:	8006;	IEF.
DR	PHCI-2DPAGE:	P10599;	-.
DR	Siena-2DPAGE:	P10599;	-.
DR	MIM;	187700;	-.
DR	InterPro:	IPR000063;	thioredoxin.
DR	pfam:	PF00085;	thioredox.
DR	PRINTS:	PR00421;	THIOREDOXIN.
DR	PROSITE:	PS00194;	THIOREDOXIN; 1.
KW	Redox-active center;	Electron transport;	3D-structure.
FT	INIT_MET	0	
FT	DISULFID	31	34 REDOX-ACTIVE (BY SIMILARITY).
FT	CONFLICT	38	K -> N (IN REF. 2).
FT	CONFLICT	73	M -> T (IN REF. 2).
FT	STRAND	2	3
FT	HELIX	7	16
FT	TURN	18	19
FT	STRAND	22	27
FT	HELIX	32	35
FT	TURN	36	37
FT	HELIX	38	41
FT	TURN	42	43
FT	HELIX	44	47
FT	STRAND	52	57
FT	TURN	58	60
FT	HELIX	62	68
FT	TURN	69	69
FT	STRAND	75	80
FT	TURN	81	82
FT	STRAND	83	89
FT	HELIX	94	103
SO	SEQUENCE	104 AA;	11606 MW; 7FFABDF3B6BE33A CRC64;

Query Match	32.9%;	Score 185;	DB 1;	Length 104;
Best Local Similarity	35.7%;	Pred. No. 3.3e-14;		
Matches 45;	Conservative 11;	Mismatches 28;	Indels 42;	Gaps 2;

OY	2	VOIQDTEFKETFLTAAGHKLAIVQFSSKRRCGPCRMPPVFHELA-----	47
Db	1	VKQIESKTAFEPALDAADGKLVDVDFSATWCGPCMKIKPEFHSLSEKYSNVITFEVDD	60
OY	48	-----TCHIKRIPTPFOMFKSKQVTLSRIKRIICVCYBSGMSNLIFFECGADAKKLEA	101
Db	61	CODVASDEEVCMPFIQGFKKQKYG-----EFSGANKEKLEA	98
OY	102	KTOELM 107	
Db	99	TINELV 104	

RESULT	2	
THIO_SHEEP	STANDARD;	PRT; 104 AA.
ID	THIO_SHEEP	
AC	P50413;	
DT	01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Thioredoxin.	
GN	TXX.	
OS	Ovis aries (Sheep).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;	
CC	Bovidae; Caprinae; Ovis.	
NB	NCBI_TaxID=9940;	
RP	[1]	SEQUENCE FROM N.A.


```

RX MEDLINE=95078463; PubMed=7987015;
RA Droogmans L., Cleuter Y., Wollman E.E., Kettmann R., Burny A.;
RT "Nucleotide sequence of ovine thioredoxin cDNA.";
RL DNA Seq. 4:277-279(1994).
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: Z25864; CAA81083.1; -.
CC HSSP: P10599; IERT.
CC InterPro: IPR000063; Thioredoxin.
CC Pfam: PF00085; thiorcd; 1.
CC PRINTS: PR00421; THIOREDOXIN.
CC PROSITE: PS00194; THIOREDOXIN; 1.
CC Redox-active center; Electron transport.
CC INIT_MET 0 0 BY SIMILARITY.
CC FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).
CC FT DISULFID 104 AA; 11711 MW; 506CE54C56A2208D CRC64;
CC SEQUENCE
SO
Query Match 32.9%; Score 185; DB 1; Length 104;
Best Local Similarity 34.9%; Pred. No. 3.3e-14;
Matches 44; Conservative 13; Mismatches 27; Indels 42; Gaps 2;
OY 2 VOIHKDNEKFTPLTAGHKLAVQFSKRCGPKRMFPVHELAE----- 47
DB 1 VKQIESYFAQFQALNSAGKLVVDFSATWCGPKMKIPFHAISEKKNVNFLEVDVD 60
OY 48 -----TCHIKTIPTPFQMFKSKQVTLFSRIKRIICYSRGSFMSNLIFFPCGADAKKLEA 101
DB 61 CKDIAAECEVACMPTFPQFFKKGQV-----EFGANKKLEA 98
OY 102 KTOELM 107
DB 99 TINELI 104
RESULT 3
THIO_RABIT STANDARD; PRT; 104 AA.
ID THIO_RABIT
AC P08628;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioredoxin.
GN TXN.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE.
RC TISSUE=Bone marrow;
RX MEDLINE=88257078; PubMed=3164311;
RA Johnson R.S., Mathews W.R., Biemann K., Hopper S.;
RT "Amino acid sequence of thioredoxin isolated from rabbit bone marrow
RT determined by tandem mass spectrometry.";
RL J. Biol. Chem. 263:9589-9597(1988).
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC PIR: A28086; A28086.
DR HSSP: P10599; IERV.

```

```

DR InterPro: IPR000063; Thioredoxin.
DR Pfam: PF00085; thiorcd; 1.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00194; THIOREDOXIN; 1.
DR Redox-active center; Electron transport.
CC FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).
CC FT DISULFID 104 AA; 11629 MW; CAB6E5EBEC231F CRC64;
CC SEQUENCE
SO
Query Match 32.7%; Score 184; DB 1; Length 104;
Best Local Similarity 34.9%; Pred. No. 4.3e-14;
Matches 44; Conservative 13; Mismatches 27; Indels 42; Gaps 2;
OY 2 VOIHKDNEKFTPLTAGHKLAVQFSKRCGPKRMFPVHELAE----- 47
DB 1 VKQIESAFQFQVDSADKLVVDFSATWCGPKMKIPFHAISEKKNVNFLEVDVD 60
OY 48 -----TCHIKTIPTPFQMFKSKQVTLFSRIKRIICYSRGSFMSNLIFFPCGADAKKLEA 101
DB 61 CKDIAAECEVACMPTFPQFFKKGQV-----EFGANKKLEA 98
OY 102 KTOELM 107
DB 99 TINELI 104
RESULT 4
THIO_BOVIN STANDARD; PRT; 104 AA.
ID THIO_BOVIN
AC O97680;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioredoxin.
GN TXN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Gotoh S., Terashima H., Yagi K., Mizoguchi T.;
RT "Cloning and sequence of bovine thioredoxin cDNA.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: AF104105; AAC83380.1; -.
CC HSSP: P10599; IERT.
CC InterPro: IPR000063; Thioredoxin.
CC Pfam: PF00085; thiorcd; 1.
CC PRINTS: PR00421; THIOREDOXIN.
CC PROSITE: PS00194; THIOREDOXIN; 1.
CC Redox-active center; Electron transport.
CC INIT_MET 0 0 BY SIMILARITY.
CC FT DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).
CC FT DISULFID 104 AA; 11681 MW; 506CF9696A2208D CRC64;
CC SEQUENCE
SO
Query Match 32.4%; Score 182; DB 1; Length 104;
Best Local Similarity 34.9%; Pred. No. 7.3e-14;
Matches 44; Conservative 12; Mismatches 28; Indels 42; Gaps 2;

```



```

QY 31.9%; Score 179; DB 1; Length 104;
Query Match Similarity 34.7%; Pred. No. 1,66-13;
Best Local Similarity
Matches 43; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

QY 2 VOI|KDTNEKFTLTAAGHKLAVVOFSSKRCGPCRMFPVPEHEL----- 45
      |:::| 1:|:::| 1:|:::| 1:|:::| 1:|:::|
Db 1 VKLIESKEAFQALNADKLVVDFSSATWCGPCMKIKPFHSLDKYSNVVLEVDVD 60

QY 46 ----AECGHIHTIPFGQFKKSOKVTLRSRIKRIICCRSGMSMLIFEGCADAKKLEA 101
      |:::| 1:|:::| 1:|:::| 1:|:::| 1:|:::|
Db 61 CQVVAADLEVACMTPFQFKKQKVG-----EFGSANKLEA 98

QY 102 KTOE 105
      |
Db 99 TITE 102

RESULT 7
THIO_MACMU STANDARD; PRT; 104 AA.
AC P29451;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioedoxin.
GN TXN.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=92181438; PubMed=1543487;
RA An G., Wu R.;
RT "Thioedoxin gene expression is transcriptionally up-regulated by
RL retinol in monkey conducting airway epithelial cells.";
RL Biochem. Biophys. Res. Commun. 183:170-175(1992).
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC .THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M84643; AAA36921.1; .
CC PIR; JS0667; JS0667.
CC HSSP; P10599; 1ERT.
CC InterPro; IPR000063; Thioedoxin.
CC Pfam; PF00085; thioed; 1.
CC PRINTS; PR00421; THIOREDOXIN.
CC PROSITE; PS00194; THIOREDOXIN; 1.
CC Redox-active center; Electron transport.
CC INTIMET 0 BY SIMILARITY.
CC DISULFET 31 34 REDOX-ACTIVE (BY SIMILARITY).
CC FT 0 BY SIMILARITY.
CC SEQUENCE 104 AA; 11606 MW; C804D5152F8870EB CnC64;

QY Query Match 31.5%; Score 177; DB 1; Length 104;
Best Local Similarity 34.9%; Pred. No. 2,7e-13;
Matches 44; Conservative 11; Mismatches 29; Indels 42; Gaps 2;

QY 2 VOI|KDTNEKFTLTAAGHKLAVVOFSSKRCGPCRMFPVPEHEL----- 47
      |:::| 1:|:::| 1:|:::| 1:|:::| 1:|:::|
Db 1 VQIISKAFAFGALDAGKLVVDFSSATWCGPCMKIKPFHSLDKYSNVVLEVDVD 60

```

ID	THIO_CHICK	STANDARD	PRT	104 AA
01	01-AUG-1988	(Rel. 08, Created)		
02	01-AUG-1988	(Rel. 08, Last sequence update)		
03	16-OCT-2001	(Rel. 40, Last annotation update)		
04	Thioredoxin.			
05	Gallus gallus (Chicken).			
06	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
07	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
08	Gallus.			
09	NCBI_TaxID=9031;			
10	SEQUENCE FROM N.A.			
11	MEDLINE=88257080; PubMed=2838473;			
12	Jones S.W., Luk K.-C.:			
13	"Isolation of a chicken thioredoxin cDNA clone. Thioredoxin mRNA is			
14	differentially expressed in normal and Rous sarcoma virus-transformed			
15	chicken embryo fibroblasts."			
16	J. Biol. Chem. 263:9607-9611(1988).			
17	-1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS			
18	THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,			
19	TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.			
20	-1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.			
21	-----			
22	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
23	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
24	the European Bioinformatics Institute. There are no restrictions on its			
25	use by non-profit institutions as long as its content is in no way			
26	modified and this statement is not removed. Usage by and for commercial			
27	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
28	or send an email to license@isb-sib.ch).			
29	-----			
30	EMBL: J03882; AAA49092.1; -			
31	PIR: A30006; A30006.			
32	HSSP; P10599; 1ERT.			
33	InterPro; IPR000063; Thioredoxin.			
34	Pfam; PF00085; thiredox.1.			
35	PRINTS; PR00421; THIOREDOXIN.			
36	PROSITE; PS00194; THIOREDOXIN; 1.			
37	Redox-active center; Electron transport.			
38	INT_MET 0			
39	DISULFID 31 34 REDOX-ACTIVE (BY SIMILARITY).			
40	SEQUENCE 104 AA: 11569 MW; 60B6B5759010BB12 CRC64;			
41	-----			
42	Query Match 28.5%; Score 160; DB 1; Length 104;			
43	Best Local Similarity 31.0%; Pred. No. 2.3e-11;			
44	Matches 39; Conservative 15; Mismatches 30; Indels 42; Gaps 2;			
45	-----			
46	2 VOIIDTNEFKFTLAAGHKILAVVOFSSKRCGPCRMPVFHEL			
47	1 KVSQGNLADFEELKAGELKLVVDSATWGPCMKIKRFFHSLDCKEGDVYFIEDVD 60			
48	46 ---AETCHIKITIPFPQFKKSQKVTLSRIKRIITICYSRSGPMSNLIEFGADAKKLEA 101			
49	61 AADVATHDVCVKMPFQFKNKGKVO-----EFGANKKELEE 98			
50	99 TIKSLV 104			

```

RESULT 9
THIO_COPCM STANDARD: PRT: 106 AA.
ID THIO_COPCM
AC 09UM02:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ThioRedoxin (Allergen Cop c 2).
OS Coprinus comatus (Shaggy mane).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Coprinaceae; Coprinus.
OX NCBI_TaxID=56187;
RN [1]
RP SEQUENCE FROM N.A.
RA Brander K.A., Cramer R., Schuermann P., Pichler W.J., Helbling A.;
RT "Coprinus thioRedoxin as inhalative allergen and crossreactive human
RT autoantigen."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ242791; CAB52130.1; -.
DR HSSP: P80028; 1TFOF.
DR InterPro: IPR000063; ThioRedoxin.
DR Pfam: PF00085; ThioRed. 1.
DR PRINTS: PR00421; THIOREDOXIN.
DR PROSITE: PS00194; THIOREDOXIN; 1.
RW Redox-active center; Electron transport; Allergen.
FT DISULFID 30 33 REDOX-ACTIVE (BY SIMILARITY).
FT SEQUENCE 106 AA; 11772 MW; 05A2155B2108C69 CRC64;
SQ

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98139906; PubMed=9473519;
RA Miranda-Vizuet A., Gustafson J.-A., Spyrou G.;
RT "Molecular cloning and expression of a cDNA encoding a human
RT thioRedoxin-like protein."
RL Biochem. Biophys. Res. Commun. 243:284-288(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Lee K.-K., Murakawa M., Takahashi S., Tsubuki S., Sakamaki K.,
RA Yonehara S.;
RT "Purification, cloning and characterization of TRP32, a novel
RT thioRedoxin-related protein of 32 kDa."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Zhou Y., Pan M.H., Yuan J.G., Qiang B.Q.;
RT "The discovery of a new gene that has high homology to the human
RT thioRedoxin gene."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20284953; PubMed=10826702;
RA Miranda-Vizuet A., Spyrou G.;
RT "Genomic structure and chromosomal localization of human
RT thioRedoxin-like protein gene (txl).";
RL DNA Seq. 10:419-424(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.22 ANGSTROMS) OF 1-108.
RA Jin J., Chen X., Guo Q., Yan J., Qiang B., Rao Z.;
RT "Crystal structure of the catalytic domain of a new human thioRedoxin-
RT like protein."
RL Submitted (NOV-2000) to the PDB data bank.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF003938; AAC39599.1; -.
DR EMBL: AF025659; AAC39898.1; -.
DR EMBL: AF051896; AAC05830.1; -.
DR EMBL: AF143897; AAF66676.1; -.
DR EMBL: AF143890; AAF66676.1; JOINED.
DR EMBL: AF143891; AAF66676.1; JOINED.
DR EMBL: AF143892; AAF66676.1; JOINED.
DR EMBL: AF143893; AAF66676.1; JOINED.
DR EMBL: AF143894; AAF66676.1; JOINED.
DR EMBL: AF143895; AAF66676.1; JOINED.
DR EMBL: AF143896; AAF66676.1; JOINED.
DR EMBL: BC001156; AAH01156.1; -.
DR PDB: 1GH2; 01-MAY-01.
DR MIM: 603049; -.
DR InterPro: IPR000063; ThioRedoxin.
DR Pfam: PF00085; ThioRed. 1.
DR PROSITE: PS00194; THIOREDOXIN; 1.
RW Redox-active center; Electron transport; 3D-structure.
FT DOMAIN 1 109 THIOREDOXIN.
FT DISULFID 34 37 REDOX-ACTIVE.
FT SEQUENCE 289 AA; 32251 MW; B2CC0BD8042225C2 CRC64;
SQ

```

```

DR   PCHI-2DPAGE; 09z7P5; -.
DR   TIGR: CP0088; -.
DR   InterPro: IPR000063; Thioresdoxin.
DR   Pfam: PF00085; Thioed; 1.
DR   PRINTS: PR00421; THIOREDOXIN.
DR   PROSITE: PS00194; THIOREDOXIN_1.
KW   Redox-active center; Electron transport; Complete proteome.
FT   DISULFID 28      REDOX-ACTIVE (BY SIMILARITY).
SQ   SEQUENCE 102 AA; 11292 MW; 76A190218324BA68 CRC64;

Query Match          19.8%; Score 111; DB 1; Length 102;
Best Local Similarity 29.6%; Pred. No. 8,1e-06;
Matches 32; Conservative 18; Mismatches 28; Indels 30; Gaps 5;

OY    1 MVGIKDTNEFKETPLTAAGHKLAVVQSSKRRCGPCKRMFPVFHEI----- 45
       ||||| | : | : | : | : | : | : | : | : | : | : | : | : |
DB    1 MKVIISSEN-PDSFI-ASG--LVLDPFPAEWCGPCRMPLPILENTLAELPHVTGINKIND 56
       - - - - - AERCHIKITPTFMFKKSKSVTLFRKRIICICRSFPMNLI 88
OY    46 -----AERCHIKITPTFMFKKSKSVTLFRKRIICICRSFPMNLI 88
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
DB    57 ENSKPATYEVSIPITLIFRDGNEVA-----RVGLKDKKEFLTNLI 98

RESULT 12
TRX2_SCHPO STANDARD: PRT; 102 AA.
AC O1463; Q9UTS9;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioresdoxin II (TR-II).
OS TRX2 OR SPACTD4.07C.
GN Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_Taxid=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21167391; PubMed=11267679;
RA Cho Y.-W., Shin Y.H., Kim Y.-T., Kim H.-G., Lee Y.-J., Park E.-H.,
RA Fuchs J.A., Lim C.-U.;
RT "Characterization and regulation of Schizosaccharomyces pombe gene
RT encoding thioresdoxin.";
RL Blochim. Biophys. Acta 1518:194-199(2001).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RA Lenaerts G., Perret E., Bonin O., Picard A., Caput D.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RA Gentiles S., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
CC TO A DISULFIDE & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
CC -!- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/sib.ch)
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF251279; AAF76881.1; -
DR EMBL: AJ0003819; CA060033.1; -
DR EMBL: Z99532; CAB16724.1; -.
DR HSSP: P80028; ITOF.
```

DR InterPro: IPR000063; Thioresdoxin.
 DR Pfam: PF00085; Thioresdoxin.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KW Redox-active center; Electron transport.
 FT INTL MET 0 BY SIMILARITY.
 FT DISULFID 29 32 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 102 AA; 11166 MW; 7069F4ACDAC34595 CRC64;

Query Match 19.6%; Score 110; DB 1; Length 102;
 Best Local Similarity 25.0%; Pred. No. 1e-05;
 Matches 30; Conservative 17; Mismatches 29; Indels 44; Gaps 3;

QY 2 VOIINDNEFKTLTAAGHKLAVVOSSKRCGPCRMPPEVFH----- 43
 ID THIF_PEA STANDARD; PRT; 102 AA.
 AC P29450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thioresdoxin I (TR-1).
 GN TRX1.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21167391; PubMed=11267679;
 RA Cho Y.-W., Shin Y.H., Kim Y.-T., Kim H.-G., Lee Y.-J., Park E.-H.,
 RA Fuchs J.A., Lim C.-J.;
 RT "Characterization and regulation of Schizosaccharomyces pombe gene
 RT encoding thioresdoxin."
 RL Biochim Biophys Acta 1518:194-199(2001).
 CC -I- FUNCTION: THIOREDOXIN PARTICIPATES IN VARIOUS REDOX REACTIONS
 CC THROUGH THE REVERSIBLE OXIDATION OF ITS ACTIVE CENTER DITHIOL,
 CC TO A DISULFIDE, & CATALYZES DITHIOL-DISULFIDE EXCHANGE REACTIONS.
 CC -I- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF192765; AAF05765.1; -
 DR InterPro: IPR000063; Thioresdoxin.
 DR Pfam: PF00085; Thioresdoxin.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KW Redox-active center; Electron transport.
 FT INTL MET 0 BY SIMILARITY.
 FT DISULFID 29 32 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 102 AA; 11307 MW; 70781D7BADB2EAB5 CRC64;

Query Match 19.2%; Score 108; DB 1; Length 102;
 Best Local Similarity 24.0%; Pred. No. 1.8e-05;
 Matches 29; Conservative 18; Mismatches 30; Indels 44; Gaps 3;

QY 2 VOIINDNEFKTLTAAGHKLAVVOSSKRCGPCRMPPEVFH----- 43
 ID THIF_PEA STANDARD; PRT; 182 AA.
 AC P29450;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thioresdoxin F-type, chloroplast precursor (TRX-F).
 DE Pisum sativum (Garden pea).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
 NX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=leaf;
 RC MEDLINE=92256804; PubMed=1581563;
 RX Lepiniec L., Hodges M., Gadal P., Cretin C.;
 RT "Isolation, characterization and nucleotide sequence of a full-length
 RT pea cDNA encoding thioresdoxin-F".
 RL Plant Mol. Biol. 18:1023-1025(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96215867; PubMed=8642611;
 RA Sahravay M., Hecht V., Lopez Jaramillo J., Chueca A., Chartier Y.,
 RA Meyer Y.;
 RT "Iron position as an evolutionary marker of thioresdoxins and
 RT thioresdoxin domains."
 RL J. Mol. Evol. 42:442-451(1996).
 CC -I- FUNCTION: PARTICIPATES IN VARIOUS REDOX REACTIONS THROUGH THE
 CC REVERSIBLE OXIDATION OF THE ACTIVE CENTER DITHIOL, TO A DISULFIDE.
 CC THE F FORM IS KNOWN TO ACTIVATE A NUMBER OF ENZYMES OF THE
 CC PHOTOSYNTHETIC CARBON CYCLE.
 CC -I- SUBCELLULAR LOCATION: Chloroplast.
 CC -I- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. PLANT F-TYPE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X63537; CAA45098.1; -
 DR EMBL: U35830; AAC49357.1; -
 DR PIR: S20929; S20929.
 DR HSSP: P10599; IAIU
 DR InterPro: IPR000063; Thioresdoxin.
 DR Pfam: PF00085; Thioresdoxin.
 DR PROSITE: PS00194; THIOREDOXIN; 1.
 KW Redox-active center; Electron transport; Chloroplast; Transit peptide.
 FT TRANSIT 1 69 CHLOROPLAST F-TYPE.
 FT CHAIN 70 182 THIOREDOXIN F-TYPE.
 FT DISULFID 106 109 REDOX-ACTIVE (BY SIMILARITY).
 SO SEQUENCE 182 AA; 19775 MW; 158FC352CB95E0F1 CRC64;

Query Match 19.0%; Score 106.5; DB 1; Length 182;
 Best Local Similarity 29.9%; Pred. No. 4.6e-05;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:25:56 ; Search time 48.76 Seconds

(without alignments)
210.861 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562

Sequence: 1 MVLKIDPNEKFTLTAAQH.....LFEFGADAKKLEAKTQELM 107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR_71:*

1: p1r1:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	190	33.8	105	1 JH0568	thioredoxin [valid
2	185	32.9	105	1 S04107	thioredoxin - mus
3	184	32.7	104	1 A28086	thioredoxin - rabb
4	184	32.7	105	1 S04352	thioredoxin - rat
5	182	32.4	105	1 JS0667	thioredoxin - rhes
6	165	29.4	105	1 A30006	thioredoxin - chic
7	118	21.0	289	2 JCS938	thioredoxin-like p
8	115	20.5	103	2 T39085	thioredoxin II - f
9	111	19.8	102	2 D72052	thioredoxin CPO088
10	111	19.8	102	2 H86572	thioredoxin f (limpor
11	106.5	19.0	182	2 S20929	thioredoxin f prec
12	103.5	18.4	190	2 S04661	thioredoxin f prec
13	103	18.3	148	2 B96721	probable thioredox
14	102.5	18.2	104	1 TXB11	thioredoxin I - ye
15	101.5	18.1	107	1 S47867	thioredoxin-like p
16	100.5	17.9	133	2 S58123	thioredoxin (clone
17	100.5	17.9	182	2 T07837	thioredoxin f prec
18	99	17.6	105	1 A46264	thioredoxin 1 - sl
19	98	17.4	104	2 C46264	thioredoxin 3 - sl
20	96.5	17.2	191	2 T12261	thioredoxin f prec
21	95.5	17.0	121	2 T39387	thioredoxin II - f
22	95	16.9	102	2 C81660	thioredoxin f prec
23	94	16.7	102	2 B71503	probable thioredox
24	94	16.7	131	2 T50865	thioredoxin-like p
25	94	16.7	131	2 T50862	thioredoxin-like p
26	94	16.7	131	2 T50862	thioredoxin-like p
27	94	16.7	281	2 S49353	protein S2 - Pala
28	94	16.7	282	2 S49352	protein S1 - Pala
29	93.5	16.6	275	2 T00710	thioredoxin homolo

30	92.5	16.5	113	1 S57775	thioredoxin h, cyt
31	91	16.2	221	2 T04271	probable thioredox
32	90.5	16.1	132	2 T04090	thioredoxin-like p
33	89	15.8	88	2 B46264	thioredoxin 2 - sl
34	89	15.8	103	1 TXB12	thioredoxin II - y
35	88.5	15.7	107	1 S31915	thioredoxin - red
36	88.5	15.7	107	2 T33843	hypothetical prote
37	88	15.7	131	2 T50863	thioredoxin-like p
38	86	15.3	118	2 S58120	thioredoxin (clone
39	85.5	15.2	106	2 A49888	thioredoxin - Penl
40	85.5	15.2	114	1 J02242	thioredoxin h - Ar
41	85.5	15.2	172	1 S38909	thioredoxin m prec
42	85.5	15.2	302	2 A96789	protein T23E18.2 f
43	85	15.1	119	2 T50867	thioredoxin-like p
44	84.5	15.0	156	2 T08086	tyrosine light chain
45	84.5	15.0	236	2 T09930	thioredoxin homolo

ALIGNMENTS

RESULT 1

JH0568

thioredoxin [validated] - human

N:Alternate names: ATL-derived factor (ADF); eosinophil cytotoxicity-enhancing factor

C:Species: Homo sapiens (man)

C:Date: 30-Jun-1992 #sequence.revision 26-May-1994 #text.change 15-Sep-2000

C:Accession: JH0568; S04106; S44375; A31993; P10079; A60749; A38922; S53453; A60870

R:Tagaya, Y.; Maeda, Y.; Mitsui, A.; Kondo, N.; Matsui, H.; Hamuro, J.; Brown, N.; Al

Gene 102, 221-228, 1991

A:Title: Isolation and characterization of human thioredoxin-encoding genes.

A:Reference number: JH0568; MUID:91340156

A:Molecule type: DNA

A:Residues: 1-105 <TAG1>

A:Cross-references: EMBL:X54539; NID:937455; PIDN:CAA38410.1; PID:9825724; EMBL:X5454

R:Tagaya, Y.; Maeda, Y.; Mitsui, A.; Kondo, N.; Matsui, H.; Hamuro, J.; Brown, N.; Al

EMBO J. 8, 757-764, 1989

A:Title: ATL-derived factor (ADF), an IL-2 receptor/Tac inducer homologous to thiore

A:Reference number: S04106; MUID:89231607

A:Molecule type: mRNA

A:Accession: S04106

A:Residues: 1-105 <TAG1>

A:Cross-references: GB:X77584; NID:9453963; PIDN:CAA54687.1; PID:9453964

A>Note: this sequence has been revised in reference S44375

EMBO J. 13, 2244, 1994

A:Reference number: S44375; MUID:94244626

A:Contents: erratum

A:Accession: S44375

A:Molecule type: mRNA

A:Residues: 1-105 <TAG2>

A:Cross-references: EMBL:X77584; NID:9453963; PIDN:CAA54687.1; PID:9453964

R:Mollman, E.E.; d'Aurilol, L.; Rimsky, L.; Shaw, A.; Jacquot, J.P.; Wingfield, P.; G

J. Biol. Chem. 263, 15506-15512, 1988

A:Title: Cloning and expression of a cDNA for human thioredoxin.

A:Reference number: A31993; MUID:89008454

A:Molecule type: mRNA

A:Accession: A31993

A:Residues: 1-38, 'N', '40-73', 'T', '75-105 <MOL>

A:Cross-references: GB:J04026; NID:9339648; PIDN:AAA74596.1; PID:9339649

R:Martin, H.; Dean, M.

Biochem. Biophys. Res. Commun. 175, 123-128, 1991

A:Title: Identification of a thioredoxin-related protein associated with plasma memb

A:Reference number: P10079; MUID:91151337

A:Molecule type: protein

A:Residues: 2-13, 'X', '15 <MAR>

R:Silberstein, D.S.; Ali, M.H.; Baker, S.L.; David, J.R.

J. Immunol. 143, 979-983, 1989

A:Title: Human eosinophil cytotoxicity-enhancing factor. Purification, physical char

A:Reference number: A60749; MUID:89309777

A:Accession: A60749

A: Molecule type: protein
A: Residues: 2-12, 'K', 14-15, 'XX', 18-19, 'X', 21-22 <STL>
A: Note: The abstract is inconsistent with figure 4 in having one undetermined residue at position 12.
J: Rimsky, L., Waksuigi, H., Ferrate, P., Robin, P., Capdevielle, J., Tursz, T., Fradelizi, J.
J: Immunol. 136, 3304-3310, 1986
A: Title: Purification to homogeneity and NH-2-terminal amino acid sequence of a novel inducible protein from human neutrophils.
A: Reference number: A38922; MUID:86169684
A: Accession: A38922
A: Molecule type: protein
A: Residues: 2-16 <WAK>
R: Dean, M.F.; Martin, H.; Sansom, P.A.
Biochem. J. 304, 861-867, 1994
A: Title: Characterization of a thioredoxin-related surface protein.
A: Reference number: S53453; MUID:95118305
A: Accession: S53453
A: Molecule type: protein
A: Residues: 1-21; 38-57 <DEA>
A: Note: described to be a surface-associated thioredoxin
R: Waksuigi, H., Rimsky, L., Mahe, Y., Kamel, A.M., Fradelizi, D., Tursz, T., Bertoglio, Proc. Natl. Acad. Sci. U.S.A. 84, 804-808, 1987
A: Title: Epstein-Barr virus-containing B-cell line produces an interleukin 1 that it uses for growth.
A: Reference number: A60870; MUID:87118252
A: Contents: annotation
R: Weichsel, A.; Gasdaska, J.R.; Powis, G.; Montfort, W.R.
Submitted to the Brookhaven Protein Data Bank, February 1996
A: Reference number: A65533; PDB:1ERT
A: Contents: annotation; X-ray crystallography, 1.7 angstroms, reduced form, residues 1-11
R: Weichsel, A.; Gasdaska, J.R.; Powis, G.; Montfort, W.R.
Submitted to the Brookhaven Protein Data Bank, February 1996
A: Reference number: A65534; PDB:1ERU
A: Contents: annotation; X-ray crystallography, 2.1 angstroms, oxidized form, residues 1-11
R: Forman-Kay, J.D.; Clore, G.M.; Gronenborn, A.M.
Submitted to the Brookhaven Protein Data Bank, December 1990
A: Reference number: A50924; PDB:4TRX
A: Contents: annotation; conformation by (1)H-NMR, residues 1-73, 'T', 75-105
R: Forman-Kay, J.D.; Clore, G.M.; Wingfield, P.T.; Gronenborn, A.M.
Biochemistry 30, 2685-2698, 1991
A: Title: High-resolution three-dimensional structure of reduced recombinant human thioredoxin.
A: Reference number: A38953; MUID:91159399
A: Contents: annotation; conformation by (1)H- and (15)N-NMR
A: Comment: This small ubiquitous protein functions in many intracellular biological pathways.
C: Genetics:
A: Gene: GDB:TXN
A: Cross-references: GDB:120475; OMIM:187700
A: Map position: 9q31-q31
A: Introns: 8/3; 43/3; 63/3; 85/3
C: Superfamily: thioredoxin; thioredoxin homology
C: Keywords: redox-active disulfide
E: 2-105/Product: thioredoxin #status experimental <MAT>
E: 9-92/Domains: thioredoxin homology <THR>
E: 32-35/Disulfide bonds: redox-active #status experimental

N:Alternate names: ATL-derived factor (ADF)
C:Species: Mus musculus (house mouse)
C:Date: 21-Nov-1993 #sequence_revision 17-Oct-1997 #text_change 11-Jun-1999
C:Accession: J04068; S44376; S04107
R:Matsui, M.; Taniguchi, Y.; Hirota, K.; Taketo, M.; Yodai, J.
Gene 152, 165-171, 1995
A:Title: Structure of the mouse thioredoxin-encoding gene and its processed pseudogen
A:Reference number: J04068; MUID:95137382
A:Accession: J04068
A:Molecule type: DNA
A:Residues: 1-105 <Mat>
A:Cross-references: DDBJ:D21855; NID:9517128
R:Tagay, Y.; Maeda, Y.; Mitsu, A.; Kando, N.; Matsui, H.; Hamuro, J.; Brown, N.; Ar
EMBO J. 13, 2244, 1994
A:Reference number: S44375; MUID:94244626
A:Contents: erratum
A:Accession: S44376
A:Molecule type: mRNA
A:Residues: 1-105 <Tr>
A:Cross-references: EMBL:X77585; NID:9453971; PIDN:CAA54688.1; PID:9453972
R:Tagay, Y.; Maeda, Y.; Mitsu, A.; Kondo, N.; Matsui, H.; Hamuro, J.; Brown, N.; Ar
EMBO J. 8, 757-764, 1989
A:Title: ATL-derived factor (ADF), an IL-2 receptor/Trac inducer homologous to thiore
A:Reference number: S04106; MUID:89251607
A:Accession: S04107
A:Molecule type: mRNA
A:Residues: 1-93, 'N', '94-96', 'ALT', '100-104', 'S' <TAG2>
A:Cross-references: GB:X77585
C:Comment: This small ubiquitous protein functions in many intracellular biological p
A:Genetics:
A:Gene: MGI:Txn
A:Cross-references: MGI:36258
A:Map position: 4:24.6
C:Introns: 29/2; 44/1; 84/2
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
R:9-32/Domain: thioredoxin homology <THR>
R:32-35/Disulfide bonds: redox-active <status predicted

```

Query Match 33.8%; Score 190; DB 1; Length 105;
Best Local Similarity 36.2%; Pred No. 1,3e-13;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2

QY 1 MWQIKIDNEFKFTPLAAGHKLAVVQSSKRCGPCKRMFVPVHELAE----- 47
   ||: | : | | | | | | | | | | | | | | | |
Db 1 MWQIESKTAFGDALDAAGDKLVVVDFSATWGGPCMKIKPFHSLSEKSNVTFLEVDVD 60
QY 48 -----TCHITITPTPFQMFKKSKQVTLFSRIKRIICCYRSGPMSNLIFFCGADAKKLE 100
   | : | | | | | | | | | | | | | | | |
Db 61 DCQADVASECEVAKCMPTFGFFKKQKRVG-----EFGANKKLE 98
QY 101 AKTOELM 107
   | | | |
Db 99 ATINELV 105

RESULT 2
S04107
thioredoxin - mouse

```

```

Query Match      32.9%; Score 185; DB 1; Length 105;
Best Local Similarity 35.2%; Pred. No. 4,5e-13;
Matches 44; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

OY      1 MWQIIKDTNEKTEFLTAAGHKLAVVQSSKCGCPCKRMFPVFHLE-----45
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       1 MYKLIESKEAFQELAAAGDKLVVDSEFATWCGCCKMKIKPFHSICDYSNVVELEYDVD 60
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY      46 -----AETCHIKTPTPEOMFKRSOKVTLFNRILRICCYRSGFSMNLIEFGCADAKLE 100
          |   |   |   |   |   |   |   |   |   |   |   |   |   |
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       61 DCGDVAAACEVKCMPTGTFPRFKKGQKV-----EISGANKEKLE 98
          |   |   |   |   |   |   |   |   |   |   |   |   |   |
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY      101 AKTQE 105
          |   |   |   |   |   |   |   |   |   |   |   |   |   |
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db       99 ASITE 103

RESULT    3
A28086
Thioredoxin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 30-Jun-1989 #sequence_revision 26-May-1994 #text_change 19-Oct-1995
C:Accession: A28086
R:Johnson, R.S.; Matthews, W.R.; Biemann, K.; Hopper, S.
J. Biol. Chem. 263, 9589-9597, 1988
A:Title: Amino acid sequence of thioredoxin isolated from rabbit bone marrow determined by complementary DNA sequencing
A:Reference number: A28086; MUID:88257078
A:Accession: A28086
A:Molecule type: protein
A:Residues: 1-104 <JOB>
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
C:;-91/Domain: thioredoxin homology <THR>
```


C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C/Accession: J05938
R:Miranda-Vizuet, A.; Gustafsson, J.A.; Spyrou, G.
Biochem. Biophys. Res. Commun. 243, 284-288, 1998
A:Title: Molecular cloning and expression of a cDNA encoding a human thioredoxin-like pr
A:Reference number: J05938; MUID:98139906
A/Accession: J05938
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-289 <MIR>
A/Cross-references: GB:AF003938; NID:g2897941; PIDN:AAC39599.1; PID:g2897942

Query Match 21.0%; Score 118; DB 2; Length 289;
Best Local Similarity 25.9%; Pred. No. 2.4e-05;
Matches 30; Conservative 17; Mismatches 27; Indels 42; Gaps 2;

QY 10 EKKFELTAAGHKLAVVQSSKRCGCKRMFPVFEHL-----AETC 49
DB 12 DQPELSCAGSHAVVKKFTMKCGPCLRIAPFSSMKRYPOAVFLEVDVHOCCTAATN 71
QY 50 HIKTPTFQMKRSQKVTLFSRIKRIICCYSGFMSNLIFFECGADAKKLEAKTQE 105
DB 72 NISATPTFLFFRNKVRID-----QYQGDVAGLEKIKQ 105

RESULT 8

T39085
thioredoxin II - fission yeast (Schizosaccharomyces pombe)
N:Alternate names: thioredoxine 2
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000
R:Gentiles, S.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A:Reference number: 221826
A/Accession: T39085
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-103 <GEN>
A/Cross-references: EMBL:Z99532; PIDN:CAB16724.1; GSPDB:GN00066; SPDB:SPACTD4.07c
R:Lenarts, G.; Perret, E.; Bonin, O.; Picard, A.; Caput, D.
submitted to the EMBL Data Library, October 1997
A:Description: TRX2, a fission yeast stress protein.
A:Reference number: 223075
A/Accession: T46567
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-103 <LEN>
A/Cross-references: EMBL:AJ003819; PIDN:CAA06033.1
A:Experimental source: strain 972(-)
C:Genetics:
A:Gene: trx2; SPACTD4.07c
A:Map position: 1
C:Superfamily: thioredoxin; thioredoxin homology

Query Match 20.5%; Score 115; DB 2; Length 103;
Best Local Similarity 25.6%; Pred. No. 2e-05;
Matches 31; Conservative 17; Mismatches 29; Indels 44; Gaps 3;

QY 1 MWQIKDTEKFTPLTAGHKLAVVQSSKRCGCKRMFPVFEH----- 43
DB 1 MWKQVDSSEKSIY--CQDKLVVDFEATWCGCKAIAFPFEGFSNTSDATFIKVDV 58
QY 44 ---ELAECHIKTPTFQMKRSQKVTLFSRIKRIICCYSGFMSNLIFFECGADAKKLE 100
DB 59 QLSIAAEAGVHAMPSFLYKNGEKIE-----EIVGANPAKLE 96
QY 101 A 101
DB 97 A 97

RESULT 9
thioredoxin CP0088 [imported] - Chlamydia pneumoniae (strains CWL029 and AR39)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C/Accession: D72052; E81614
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A/Accession: D72052
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-102 <ARN>
A/Cross-references: GB:AE001648; GB:AE001363; NID:g4376946; PIDN:AAD18798.1; PID:g437

A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A/Accession: E81614
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-102 <REA>
A/Cross-references: GB:AE002172; GB:AE002161; NID:g7189018; PIDN:AAF37973.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: trxA; CP0088
C:Superfamily: thioredoxin; thioredoxin homology

Query Match 19.8%; Score 111; DB 2; Length 102;
Best Local Similarity 29.6%; Pred. No. 5.4e-05;
Matches 32; Conservative 18; Mismatches 28; Indels 30; Gaps 5;

QY 1 MWQIKDTEKFTPLTAGHKLAVVQSSKRCGCKRMFPVFEHL----- 45
DB 1 MWKTISSBN-PDSFL-ASG--LVLVDFEAWCGCKRMFLPTEMLAELPHVTIGKINID 56
QY 46 -----AETCHIKTPTFQMKRSQKVTLFSRIKRIICCYSGFMSNLI 88
DB 57 ENSKPAETFEVSSIPTLILFRDGNV-----RVYGLKDKREFLTNLI 98

RESULT 10

H86572
thioredoxin [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C/Accession: H86572
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A/Accession: H86572
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-102 <STO>
A/Cross-references: GB:BA000008; NID:g8979031; PIDN:BAA98866.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: trxA
C:Superfamily: thioredoxin; thioredoxin homology

Query Match 19.8%; Score 111; DB 2; Length 102;
Best Local Similarity 29.6%; Pred. No. 5.4e-05;
Matches 32; Conservative 18; Mismatches 28; Indels 30; Gaps 5;

QY 1 MWQIKDTEKFTPLTAGHKLAVVQSSKRCGCKRMFPVFEHL----- 45

[illegible]

A:Reference number: A39847; MUID:91225027
A:Accession: B39847
A:Molecule type: DNA
A:Residues: 1-104 <MDL>
A:Cross-references: GB:M62648; NID:9173049; PIDN:AAA5178.1; PID:9173050
R:Hall, D.E.; Balesien, A.; Holmgren, A.; Reichard, P.
Eur. J. Biochem. 23, 328-335, 1971
A:Title: Yeast thioredoxin. Amino acid sequence around the active-center disulfide of th
A:Reference number: S05793; MUID:72100583
A:Accession: S05793
A:Molecule type: protein
A:Residues: 2:27-43:98-104 <HAL>
A:Note: the sequence from the summary and from Fig. 5 is inconsistent with that from pag
R:Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the EMBL Data Library, April 1995
A:Description: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII
es, of the human.
A:Reference number: S53922
A:Accession: S53922
A:Molecule type: DNA
A:Residues: 1-104 <GDE>
A:Cross-references: EMBL:Z49133; NID:9790489; PIDN:CAA89002.1; PID:9790500
A:Experimental source: strain S286C
R:Song, J.M.; Cheung, E.; Rabinowitz, J.C.
submitted to the EMBL Data Library, November 1995
A:Description: Analysis of the 15.6-kb fragment encompassing the *ADP3* gene.
A:Reference number: S61947
A:Accession: S61947
A:Molecule type: DNA
A:Residues: 1-104 <SON>
A:Cross-references: EMBL:U04083; NID:91165213; PIDN:AAA85584.1; PID:91165214
A:Experimental source: strain GPF88
R:Guerrero, P.; Barreiros, T.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64517
A:Accession: S64517
A:Molecule type: DNA
A:Residues: 1-104 <GWN>
A:Cross-references: EMBL:Z72994; NID:91333374; PIDN:CAA97236.1; PID:91323375; GSPDB:GNOC
A:Experimental source: strain S286C
R:Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousa
Yeast 12, 273-280, 1996
A:Title: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII reveal
terial electron-transferring flavoproteins (beta-chain) and of the *Escherichia coli* phos
A:Reference number: S63848; MUID:97060019
A:Accession: S63848
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-104 <GUP>
A:Cross-references: EMBL:Z49133; NID:9790489; PIDN:CAA89002.1; PID:9790500
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:
A:Gene: SGD:TRX2; TR-1; MIPS:YGR209C
A:Cross-references: SGD:S0003441; MIPS:YGR209C
A:Map position: 7R
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
F:2-104/Product: thioredoxin I #status experimental <MAT>
F:9-91/Domain: thioredoxin homology <THR>
F:31-34/Disulfide bonds: redox-active #status experimental

RESULT 15
S47867
thioredoxin-like protein - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S47867
R:Salz, H.R.; Flickinger, T.W.; Mittendorf, E.; Pelliscena-Palle, A.; Petschek, J.P.;
Genetics 136, 1075-1086, 1994
A:Title: The *Drosophila* maternal effect locus deadhead encodes a thioredoxin homolog
A:Reference number: S47867; MUID:94274010
A:Accession: S47867
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <SAL>
A:Cross-references: EMBL:L27072; NID:9435591; PIDN:AAA28937.1; PID:9435963
C:Genetics:
A:Gene: flybase:dhd
A:Cross-references: flybase:FBgn0011761
C:Superfamily: thioredoxin; thioredoxin homology
C:Keywords: redox-active disulfide
F:8-92/Domain: thioredoxin homology <THR>
F:31-34/Disulfide bonds: redox-active #status predicted

Query Match 18.1%; Score 101.5; DB 1; Length 107;
Best Local Similarity 22.4%; Pred. No. 0.00062;
Matches 26; Conservative 17; Mismatches 30; Indels 43; Gaps 2;

OY 5 IKDNEKFTFLTAGHKLAVYQFSKRCGPKRM-----FPYEH 43
Db 4 VRTMNDYHKRIEADDDKLVDFYATWCGPCKEMESTYKSLARRYSSKAVVLIKIDVDKFE 63

OY 44 ELAETCHIKTIPTFQMPFKSOKVTLFSRIKRIICCYRSGFMSNLIFFCGADAKKL 99
Db 64 ELTERYKYSMPTFYFLKQNRRLA-----SPAGADEKRL 97

Search completed: June 8, 2002, 04:29:51
Job time: 3835 sec

Query Match 18.2%; Score 102.5; DB 1; Length 104;
Best Local Similarity 28.7%; Pred. No. 0.00047;
Matches 25; Conservative 15; Mismatches 26; Indels 21; Gaps 2;

OY 1 MWIITKDNNEKFTFLTAGHKLAVYQFSKRCGPKRMFPVHELAE----- 47
Db 1 MVTQKKSASEYDSAL-ASGDKLVYVDFPRTWCGPCKMTAPMEKRAEQYSDAATFKLDVD 59

OY 48 -----TCHIKTIPTFQMPFKRSOKVT 67
Db 60 EVSDVAQAKAEVSMPTLIFYGKGKEVT 86

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 03:21:11 ; Search time 32.08 seconds
(without alignments)
81.469 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562
Sequence: 1 MWQIKDKNEKFTLLANGH.....IFEFGADAKKLEAKTOELM 107

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents_AA:*

1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*

2: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*

3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*

4: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*

5: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*

6: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	33.8	210	US-08-180-371-6	Sequence 6, Appl 1
2	190	33.8	210	PCT-US92-05707-6	Sequence 6, Appl 1
3	188	33.5	105	US-08-826-910-3	Sequence 3, Appl 1
4	185	32.9	104	5210073-1	Patent No. 5210073
5	185	32.9	105	US-08-826-910-4	Sequence 4, Appl 1
6	118	21.0	130	US-08-826-910-1	Sequence 1, Appl 1
7	85.5	15.2	106	US-08-386-729A-10	Sequence 10, Appl 1
8	84.5	15.0	111	US-09-264-419C-6	Sequence 6, Appl 1
9	77	13.7	1285	US-08-974-549A-600	Sequence 600, Appl 1
10	75	13.3	521	US-08-557-122A-32	Sequence 32, Appl 1
11	75	13.3	521	US-09-262-666-32	Sequence 32, Appl 1
12	74	13.2	129	US-08-557-122A-11	Sequence 11, Appl 1
13	74	13.2	129	US-09-262-666-11	Sequence 11, Appl 1
14	74	13.2	177	US-09-272-342B-6	Sequence 6, Appl 1
15	74	13.2	485	US-08-068-995A-1	Sequence 1, Appl 1
16	74	13.2	485	US-08-464-365-1	Sequence 1, Appl 1
17	74	13.2	505	US-08-068-995A-3	Sequence 3, Appl 1
18	74	13.2	505	US-08-464-365-3	Sequence 3, Appl 1
19	73	13.0	131	US-08-557-122A-10	Sequence 10, Appl 1
20	73	13.0	131	US-09-262-666-10	Sequence 10, Appl 1
21	73	13.0	141	US-08-557-122A-9	Sequence 9, Appl 1
22	73	13.0	141	US-09-262-666-9	Sequence 9, Appl 1
23	73	13.0	143	US-08-557-122A-8	Sequence 8, Appl 1
24	73	13.0	143	US-09-262-666-8	Sequence 8, Appl 1
25	73	13.0	163	US-08-557-122A-7	Sequence 7, Appl 1
26	73	13.0	163	US-09-262-666-7	Sequence 7, Appl 1
27	73	13.0	174	US-08-557-122A-6	Sequence 6, Appl 1

28	73	13.0	174	US-09-262-666-6	Sequence 6, Appl 1
29	73	13.0	200	US-08-557-122A-12	Sequence 12, Appl 1
30	73	13.0	200	US-09-262-666-12	Sequence 12, Appl 1
31	73	13.0	281	US-08-557-122A-5	Sequence 5, Appl 1
32	73	13.0	281	US-09-262-666-5	Sequence 5, Appl 1
33	73	13.0	511	US-08-557-122A-4	Sequence 4, Appl 1
34	73	13.0	511	US-09-262-666-4	Sequence 4, Appl 1
35	73	13.0	515	US-08-557-122A-3	Sequence 3, Appl 1
36	73	13.0	515	US-08-557-122A-34	Sequence 34, Appl 1
37	73	13.0	515	US-09-262-666-3	Sequence 3, Appl 1
38	73	13.0	515	US-09-262-666-34	Sequence 34, Appl 1
39	73	13.0	530	US-08-557-122A-35	Sequence 35, Appl 1
40	73	13.0	530	US-09-262-666-35	Sequence 35, Appl 1
41	72.5	12.9	109	US-07-745-382-22	Sequence 22, Appl 1
42	72.5	12.9	109	US-07-921-848-22	Sequence 22, Appl 1
43	72.5	12.9	109	US-08-165-301A-22	Sequence 22, Appl 1
44	72.5	12.9	109	US-08-810-436-22	Sequence 22, Appl 1
45	72.5	12.9	109	US-09-485-885-17	Sequence 17, Appl 1

ALIGNMENTS

RESULT 1
US-08-180-371-6
; Sequence 6, Application US/08180371
; Patent No. 6254861
; GENERAL INFORMATION:
; APPLICANT: Choudhury, Chandra
; TITLE OF INVENTION: Hematopoietic Growth Factor Derived
; TITLE OF INVENTION: from T Lymphocytes and Methods of Use Therefor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/180, 371
; FILING DATE: 12-JAN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 906866
; FILING DATE: 01 July 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27620/31668
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25 3856
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-180-371-6

Query Match 33.8%; Score 190; DB 4; Length 210;
Best Local Similarity 36.2%; Pred. No. 1.5e-17;
Matches 46; Conservative 11; Mismatches 28; Indels 42; Gaps 2;

CLONE:
US-08-826-910-4

QY 10 EFFETLTAAGHKLA VQSSKRCGPCKRMFE

OV 10 EKFTELTAAGHKILAVVOFSSKRCGPCCKRMFE

OV 10 EFKTFLTAAGHKILAVVOFSSKRCGPCRMEF


```

PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 600:
SEQUENCE CHARACTERISTICS:
LENGTH: 1285 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..1285
OTHER INFORMATION: /note= "fusion protein composed of enterokinase cleavable, His tagged
OTHER INFORMATION: thioresdoxin moiety and full length hTTR"
US-08-974-549A-600

```

```

Query Match          13.7%; Score 77; DB 4; Length 1285;
Best Local Similarity 23.9%; Pred. No. 0.2;
Matches 27; Conservative 15; Mismatches 33; Indels 38; Gaps 4.

QY      24  VQESSKRCGCKRMFPPEHLEATCH-----IKTIPTQMEFK 62
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      25  LVDFAHNCGCKMIAPILEIDETADYQCKLYAKLRIDHNPGTARKYGRIGIPTLLFFKN 84

QY      63  SQ----KYLFSR--IKRIICCRSGFMSN-----LIFFCGADARK 98
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      85  GEVAATKVGALSKGQLKEFLDANLAGSGSDDDKVPMHLEIIFEPAASTOR 137

RESULT 10
US-08-557-122A-32
; Sequence 32, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mailand
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase

```

NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-357-122A-32

```

Query March          13.3%; Score 75; DB 2; Length 521;
Best Local Similarity 25.6%; Pred. No. 0.11;
Matches 20; Conservative 12; Mismatches 24; Gaps 2

QY      8 TNEFTETLTAAGHKLAVVOCFSRKGCBKMPFVFHELAETCHIKRT----- 54
       | : | : | : | : | : | : | : | : | : | : | : | : |
Db     39 TDSTNEYIGS--HDVLVAEEFPAPMGCGCKMMAGEYKAAMETLVENKITLAQIDCTENDL 96

QY      55 -----PRPQMFKKS 63
       | : : : | | |
Db     97 CMEHNIPGFPSLKIFKNSS 114

RESULT 11
US-09-262-666-32
: Sequence 32, Application US/09262666
: Patent NO. 6346244
: GENERAL INFORMATION:
: APPLICANT: Hjort, Carsten Malland
: TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6346244 No. 6346244disk Of No. 6346244th America, Inc
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/262,666
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/557,122

```

; RECORD NO: 020754
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: YU, YEON-GYU

APPLICANT: KIM, SUNG-HOU
APPLICANT: RYU, JAE-REON
TITLE OF INVENTION: METHOD FOR DETECTING A SUBSTANCE HAVING AN ACTIVITY TO
TITLE OF INVENTION: INHIBIT HIV INFECTION USING IMMUNOASSAY AND VARIANT
TITLE OF INVENTION: PROTEIN USED FOR SAID METHOD
FILE REFERENCE: 2901-0125-0
CURRENT APPLICATION NUMBER: US/09/272,342B
CURRENT FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 177
TYPE: PRT
ORGANISM: homo sapiens and HIV hybrid
US-09-272-342B-6

Query Match 13.2%; Score 74; DB 4; Length 177;
Best Local Similarity 24.5%; Pred. No. 0.038;
Matches 27; Conservative 15; Mismatches 34; Indels 34; Gaps 4;

QY 24 VQFSSKRCGPCRMFPVHFLAETCH-----IKTIPFQMFKK 62
DB 25 LVDFMAEMCGPCRMIAPIIDEIADYQGLTVAKLNIDDPGAPKXYGIPFLFLFKN 84
QY 63 SQ-----KVLFSR--IKRIICYRSGFMSNLIFFCGADAKKLEAKTQEL 106
DB 85 GEVAATKRVGALSKQLKEFL-----DANLASSGSGDDDDKVPGLDEL 127

RESULT 15

US-08-068-395A-1
Sequence 1, Application US/08068395A
Patent No. 5496719

GENERAL INFORMATION:
APPLICANT: YAMADA, YUKIO
APPLICANT: ASAMI, OSAMU
APPLICANT: SUGIYAMA, HIDEHIKO
APPLICANT: IDEKOB, CHIE
APPLICANT: HOSHINO, FUMHIKO
APPLICANT: HIRAI, MASANA
APPLICANT: KAJINO, TSUTOMU
APPLICANT: IMAEDA, TAKAO
APPLICANT: SARAI, KIYOKO
TITLE OF INVENTION: POLYPEPTIDE POSSESSING PROTEIN DISULFIDE
TITLE OF INVENTION: ISOMERASE ACTIVITY GENE ENCODING THE SAME AND PROCESS FOR
TITLE OF INVENTION: PRODUCING THE SAME
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/068,395A
FILING DATE: 19930527
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-135254
FILING DATE: 27-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-44013
FILING DATE: 04-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-44014

FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5496719man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 68-228-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 485 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-068-395A-1

Query Match 13.2%; Score 74; DB 1; Length 485;
Best Local Similarity 25.8%; Pred. No. 0.14;
Matches 23; Conservative 13; Mismatches 27; Indels 26; Gaps 3;

QY 1 MVQIINKDTNEKTFLLTAGHKLAVVQFSSKRCGPCRMFPVHFLAETCHIKIT----- 54
DB 3 VVQLKKDT--FDFIKT--NDLVLAFFFAWPCGCKALAEYEBAAATLKEKNIKLAKVD 58
QY 55 -----PTFQMFKKSKQVT 67
DB 59 CTEETDLCOQHGVGEGYPTLVKFRGLNVS 87

Search completed: June 8, 2002, 04:28:49
Job time: 4058 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2002, 04:28:51 ; Search time 89.36 Seconds
(without alignments)
207.145 Million cell updates/sec

Title: US-09-746-783-19

Perfect score: 562

Sequence: 1 MVQIKDNEFKFLTAAGH.....IPEFGADAKKLEAKTQELM 107

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhcc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirts:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	313	55.7	127	11	09CQ96 mus musculus
2	202	35.9	84	4	060744 mus sapien
3	188	33.5	105	4	096K13 homo sapien
4	187	33.3	105	6	095J99 sus scrofa
5	185	32.9	105	6	097508 equus caball
6	185	32.9	105	6	09BDJ3 gallitrix
7	182	32.4	105	11	09PDR0 mus musculus
8	179	31.9	105	13	098TX1 ophiophagus
9	165	29.4	112	4	09H018 homo sapien
10	160	28.5	486	4	096RX3 aedes aegypt
11	146	26.0	107	13	09DGI3 ictalurus p
12	135.5	24.1	106	5	0963R3 aedes aegypt
13	134	23.8	107	5	09NRP3 drosophila
14	133.5	23.6	107	5	09NG21 anopheles g
15	132.5	23.6	104	5	0962B7 branchiosto
16	124	22.1	284	5	09N2K6 caenorhabd1

17	123.5	22.0	106	5	09U515 manduca sex
18	118	21.0	289	11	070379 mus musculus
19	118	21.0	289	11	0920J4 rattus norv
20	117.5	20.9	106	5	09V429 drosophila
21	115.5	20.6	104	5	09N1R2 plasmodium
22	114	20.3	106	5	096952 geodia cydo
23	110.5	19.7	104	5	09NFK9 plasmodium
24	105.5	18.8	114	5	0955W4 drosophila
25	103	18.3	148	10	09CAS1 arabisdopsis
26	102.5	18.2	101	10	022031 cyanidium c
27	102.5	18.2	106	5	0908F3 schistosoma
28	99.5	17.7	245	10	09XFI1 arabidopsis
29	98.5	17.5	178	10	09M8R5 arabidopsis
30	97	17.3	385	10	093X83 vitis labru
31	96	17.1	140	17	096YQ0 sulfolobus
32	95.5	17.0	121	3	094504 schizosacch
33	94	16.7	131	10	09S753 phalaris co
34	94	16.7	131	10	09SWG6 hordeum bul
35	94	16.7	131	10	09SWG4 lolium pere
36	93.5	16.6	98	5	025549 naegleria f
37	93.5	16.6	275	10	064654 arabidopsis
38	91.5	16.3	304	5	09N357 caenorhabd1
39	91	16.2	215	10	09XFI10 arabidopsis
40	91	16.2	221	10	09S2I7 arabidopsis
41	89	15.8	125	10	09LIX4 triticum ae
42	89	15.8	130	10	064395 triticum tu
43	89	15.8	140	10	09C9Y6 arabidopsis
44	88.5	15.7	107	5	09TX18 caenorhabd1
45	88.5	15.7	111	5	09W4D6 drosophila

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	127 AA.
09CQ96	09CQ96	09CQ96:		
AC	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	4930429J24RIK PROTEIN.			
GN	4930429J24RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
ID	09CQ96			
AC	09CQ96:			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS; MEDLINE=21085660; PubMed=11217851;			
RX				
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberters P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK015240; BAB29760.1; -			
DR	EMBL; AK006683; BAB24702.1; -			

.RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.
 .RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.
 .RT "The genome sequence of *Drosophila melanogaster*."
 .RL Science 287:2185-2195(2000).
 .RN [2]
 .RP SEQUENCE FROM N.A.
 .RA Miranda-Vizuette A.:
 .RT "Drosophila melanogaster homolog of human txi gene."
 .RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 .DR EMBL; AE003365; AAF65075.1; -
 .DR EMBL; AF143404; AAF66635.1; -
 .DR HSSP: O43396; 1GH2.
 .DR FlyBase; FBgn0035631; Txi1.
 .DR InterPro: IPR000362; Fumarate_lyase.
 .DR InterPro: IPR000063; Thired.
 .DR Pfam: PF00085; Thireded_1.
 .DR PRINTS: PR00421; THIOREDOXIN.
 .DR PROSITE: PS00163; FUMARATE_LYASES; UNKNOWN_1.
 .DR PROSITE: PS00194; THIOREDOXIN; 1.
 .KW Redox-active center.
 .FT CONFLICT 149 150 LA -> WP (IN REF. 2).
 .FT CONFLICT 177 177 H -> O (IN REF. 2).
 .SQ SEQUENCE 287 AA; 31728 MM; BB316D9D91A9432A6 CRC64;

Query Match	23.8%;	Score 134;	DB 5;	Length 287;
Best Local Similarity	27.4%;	Pred. No. 1.8e-08;		
Matches 34;	Conservative 17;	Mismatches 31;	Indels 42;	Gaps 4;

```

QY 2 VQLIANDINERIFETLLAAGHGLATLVVDESSKRCPCRKRF PVFHEH-----AET 48
D 3 VRVINDSEHFQAEALAAAGIOLVVVDFTASWCGPCRKRIAPIFEFEPTPYPPAIFLKYDVDK 62
D 49 CH-----IKTIPTFQMEFKSKQVTLTFSRIKRIITICYSRQSPMSNLIFEGGADAKLEA 101
D 63 CQDPAAGGVSAMPTFLFYRRRTNI-----DRVQ-----GADVNGLEA 100
QY 102 KTOE 105
D 101 KIOE 104

```

RESULT	14	
09NGZ1		
ID	09NGZ1	PRELIMINARY; PRT; 107 AA.
AC	09NGZ1;	
DT	01-OCT-2000 (TREMblrel. 15, Created)	
DT	01-OCT-2000 (TREMblrel. 15, Last sequence update)	
DT	01-DEC-2001 (TREMblrel. 19, Last annotation update)	
DE	THIOREDOXIN 1.	
OS	Anopheles gambiae (African malaria mosquito).	
OC	Eukaryota; Metazoa; Arthropoda; Insecta;	
OC	Pharyngota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae.	
OC	Anopheles.	
OX	NCBI_Taxid=7165;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Kanzok S., Schlimmer R.H., Becker K.;	
RT	"Thioredoxin 1 of the malaria vector Anopheles gambiae.";	
RL	Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.	
DR	EMBL: AF236124; AAF68382.1; -	
DR	HSSP: P10599; 1ERY.	
DR	InterPro: IPR000063; ThioRed.	
DR	Pfam: PF00085; thioRed. 1.	
DR	PRINTS: PR00421; THIOREDOXIN.	
DR	PROSITE: PS00194; THIOREDOXIN; 1.	
KW	Redox-active center.	
Q	SEQUENCE 107 AA; 12085 MW; 888E404B1BF9A0E7 CRC64;	

Query Match	23.8%;	Score 133.5;	DB 5;	Length 107;
Best Local Similarity	27.8%;	Pred. No. 7.3e-09;		
Matches 35; Conservative	18;	Mismatches 30;	Indels 43;	Gaps 2

```

QY 1 MVLQITNEKTEFLAAGHKLAVVQFSKRGCPCKRMFPVF-----42
Db 1 MYVWKVSEDFENKLELAADQDQVAVDFPATWCGPCVYAPLFEFQNKYADKLIVVVKVDV
QY 43 ---HELAETCHIKITPTFQMFKKSQKVTLSRIKRLICVRSQSMNLIFPCGADAKTL 99
Db 61 DECEELAAQINIASMPTFLFIRKE-----VQGFSGANAEXL 98
QY 100 EAKTQE 105
Db 99 ENFIQQ 104

```

RESULT	15	
0962B7		
ID	0962B7	PRELIMINARY; PRT: 104 AA.
AC	Q962B7;	
DT	01-DEC-2001 (TREMBLrel. 19, Created)	
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	THIOREDOXIN.	
OS	Branchiostoma belcheri (Amphioxus).	
OC	Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;	
OC	Branchiostoma.	
OX	NCBI_taxid=7741;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Huang X., Wang X., Wang X., Li L., Han H., Yang H., Zhang H.;	
RT	"Cloning of thioredoxin gene in amphioxus."	
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL: AT037884; AAK72483.1; -	
DQ	SEQUENCE 104 AA; 11431 MW; 07D453D14FC684A1 CRC64;	

```
Query Match      23.6%; Score 132.5; DB 5; Length 104;  
Best Local Similarity 31.1%; Pred. NO. 9.5e-09;  
Matches 37; Conservative 11; Mismatches 28; Indels 43; Gaps 4.  
  
QY      1 MWGIITKDNNEKTFLTAGHKLAVVQFSRRSGPCKRMEPVYHELAE----- 47  
       |||||: | | | | | : |||||: || | : || |  
Db      1 MWQMLETKAADFCLKL-AETDKLIIVDFIASWCGPCRMRTAPFEKELAEENPDVVAYKVDVD 59  
  
QY      48 -----TCHKITPTPFOMFKKSQKVTLFSSRIKRILICCRYSRGSMNLJFEFGADAKKL 99  
       || | : || | : || | : || | : || | : || |  
Db      60 ANDETACGCGISAMPTFH-----CYRNCAK--VEELCGASEEKL 96
```

Search completed: June 8, 2002, 04:36:57
Job time: 486 sec

Sat Jun 8 10:14:14 2002

us-09-746-783-19.rsp

Page 7

